

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/12, 15/19, 15/57, 15/62, C07K 16/18, 16/28, 16/24, 16/40, C12Q 1/68, G01N 33/566, 33/68, A61K 38/17, 38/19, 38/48	A2	(11) International Publication Number: WO 00/01817 (43) International Publication Date: 13 January 2000 (13.01.00)																	
(21) International Application Number: PCT/US99/12366 (22) International Filing Date: 6 July 1999 (06.07.99) (30) Priority Data: <table border="0"><tr><td>09/110,938</td><td>6 July 1998 (06.07.98)</td><td>US</td></tr><tr><td>09/114,466</td><td>13 July 1998 (13.07.98)</td><td>US</td></tr><tr><td>60/093,897</td><td>23 July 1998 (23.07.98)</td><td>US</td></tr><tr><td>09/132,968</td><td>12 August 1998 (12.08.98)</td><td>US</td></tr><tr><td>09/136,214</td><td>18 August 1998 (18.08.98)</td><td>US</td></tr><tr><td>60/099,999</td><td>11 September 1998 (11.09.98)</td><td>US</td></tr></table> (71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US). (72) Inventors: BATES, Elizabeth, Esther, Mary; 4, place Gabriel Rambaud, F-69001 Lyon (FR). LEBECQUE, Serge, J., E.; 514, Chemin du Marand, F-69380 Civrieux d'Azergue (FR). MURPHY, Erin, E.; 180 Emerson Street, Palo Alto, CA 94301 (US). MATTSON, Jeanine, D.; 559 Alvarado Street, San Francisco, CA 94114 (US). GORMAN, Daniel, M.; 6371 Central Avenue, Newark, CA 94560 (US). HEDRICK, Joseph, A.; 52-08 Quail Ridge Drive, Plainsboro, NJ 08536 (US). WANG, Luquan; 21 Hollis Road, East Brunswick,	09/110,938	6 July 1998 (06.07.98)	US	09/114,466	13 July 1998 (13.07.98)	US	60/093,897	23 July 1998 (23.07.98)	US	09/132,968	12 August 1998 (12.08.98)	US	09/136,214	18 August 1998 (18.08.98)	US	60/099,999	11 September 1998 (11.09.98)	US	NJ 08816 (US). ZLOTNIK, Albert; 507 Alger Drive, Palo Alto, CA 94306 (US). MURGOLO, Nicholas, J.; 99 Rolling Hill Drive, Millington, NJ 07946 (US). GREENE, Jonathan, R.; 457 Tillou Road, South Orange, NJ 07079 (US). JOHNSTON, James, A.; 205 Mary Alice Drive, Los Gatos, CA 95032 (US). BAZAN, Jose, Fernando; 775 University Drive, Menlo Park, CA 94025 (US). MAHONY, Daniel; 330 East 39th Street #21-A, New York, NY 10016 (US). LEES, Emma, M.; 3107 Washington Street, San Francisco, CA 94115 (US). (74) Agents: THAMPOE, Immac, J. et al.; Schering-Plough Corporation, Patent Dept., K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US). (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GD, GE, HR, HU, ID, IL, IN, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LU, LV, MD, MG, MK, MN, MX, NO, NZ, PL, PT, RO, RU, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
09/110,938	6 July 1998 (06.07.98)	US																	
09/114,466	13 July 1998 (13.07.98)	US																	
60/093,897	23 July 1998 (23.07.98)	US																	
09/132,968	12 August 1998 (12.08.98)	US																	
09/136,214	18 August 1998 (18.08.98)	US																	
60/099,999	11 September 1998 (11.09.98)	US																	
(54) Title: MAMMALIAN GENES; DENDRITIC CELL PROSTAGLANDIN-LIKE TRANSPONDER (DC-PGT), HDTEA84, HSLJD37R AND RANKL, HCC5 CHEMOKINE, DEUBIQUITINATING 11 AND 12 (DUB11, DUB12), MD-1, MD2 AND CYCLIN E2, RELATED REAGENTS AND METHODS (57) Abstract <p>Purified genes from a mammal, reagents related thereto including purified proteins, specific antibodies, and nucleic acids encoding the polypeptides are provided. Methods of using said reagents and diagnostic kits are also provided. Characterization of genes and products relating to DC-PGT (Dendritic cell prostaglandin-like transporter), HDTEA84, HSLJD37R and RANKL (related to TNF receptor family), HCC5 chemokine, Dub 11 and Dub 12 (Deubiquitinating 11 and 12), MD-1 and MD-2 (proteins which exhibit properties of ligands for proteins exhibiting a leucine-rich protein motif (LRR)) and cyclin E2.</p>																			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

MAMMALIAN GENES: DENDRITIC CELL PROSTAGLANDIN-LIKE TRANSPONDER (DC-PGT), HDTEA84, HSLJD37R AND RANKL, HCC5 CHEMOKINE, DEUBIQUITINATING 11 AND 12 (DUB11, DUB12), MD-1, MD2 AND CYCLIN E2, RELATED REAGENTS AND METHODS

FIELD OF THE INVENTION

5 The present invention pertains to compositions related to proteins which: function in cellular physiology, development, and differentiation of mammalian cells; exhibit sequence similarity to TNF receptors which function in controlling activation and expansion of mammalian cells, e.g., cells of a mammalian immune
10 system; or function in controlling the cell cycle and growth. In particular, it provides purified genes, proteins, antibodies, and related reagents useful, e.g., to separate or identify particular cell types, or to regulate activation, development, differentiation, and function of various cell types, including
15 hematopoietic cells; which exhibit high structural similarity to proteins that exhibit the biological capacity to serve as a carrier mediated transporters of charged organic anions across cellular membranes, which typically can be used in prostaglandin and thromboxane physiology, e.g., transportation, influx, efflux,
20 clearance, or degradation; which regulate or evidence development, differentiation, and function of various cell types, including hematopoietic cells; or to regulate cell division and proliferation of various cell types, including tumor cells.

25 BACKGROUND OF THE INVENTION

Prostaglandins (PGs) and thromboxanes (TXs) play widespread physiological, and therapeutic roles in health and disease such as glaucoma; pregnancy, labor, delivery, and abortion; gastric
30 protection and peptic ulcer formation; intestinal fluid secretion; liver protection and damage; airway resistance and asthma; blood pressure control; and modulation of inflammatory cells.

PGs are charged anions at physiological pH that diffuse poorly across biological membranes. This limited simple diffusion
35 appears to be augmented by carrier mediated transport in many diverse tissues such as the lung, choroid plexus, liver, anterior chamber of the eye, vagina, uterus, and placenta.

Understanding the role of prostaglandins in the development and functioning of the immune system is presently incomplete. Specifically, the influence of prostaglandins (PGs) on antigen presenting cells (APCs) of the immune system (e.g., dendritic cells) is, as yet, poorly understood.

Dendritic cells (DCs) are the most potent of antigen presenting cells. See, e.g., Paul (ed. 1993) Fundamental Immunology 3d ed., Raven Press, NY. DCs are highly responsive to inflammatory stimuli such as bacterial lipopolysaccharides (LPS) and cytokines such as tumor necrosis factor alpha (TNF α). The presence of cytokines and LPS can induce a series of phenotypic and functional changes in DC that are collectively referred to as maturation. See, e.g., Banchereau and Schmitt Dendritic Cells in Fundamental and Clinical Immunology Plenum Press, NY.

Maturation changes in DCs include, e.g., silencing of antigen uptake by endocytosis, upregulation of surface molecules related to T cell activation, and active production of a number of cytokines including TNF α and IL-12. Upon local accumulation of TNF α , DCs migrate to the T cell areas of secondary lymphoid organs to activate antigen specific T cells.

Recent data indicate that DCs secrete PGs. See, e.g., Cormann, et al. (1986) Ann. Inst. Pasteur 137:369-382. Furthermore, PGE₂ has been shown to have an influence on DC maturity and the production of cytokines by DCs. See e.g., Kalinski, et al. (1997) J. Immunol. 159:28-35; Kuhn, et al. (1997) Eur. J. Immunol. 27:3135-3142; and Rieser, et al. (1997) J. Exp. Med. 186:1603-1608.

Currently, a need exists to understand the manner in which PGs influence cells of the immune system. It seems likely that PGs, like cytokines, effect immune system development and activation. The present invention contributes to satisfying that need and is directed generally to a novel mammalian gene encoding a prostaglandin-like transporter (PGT).

In other aspects, the activation of resting T cells is critical to most immune responses and allows these cells to exert

their regulatory or effector capabilities. See, e.g., Paul (ed. 1993) Fundamental Immunology 3d ed., Raven Press, N.Y. Increased adhesion between T cells and antigen presenting cells (APC) or other forms of primary stimuli, e.g., immobilized monoclonal antibodies (mAb), can potentiate the T-cell receptor signals. T-cell activation and T cell expansion depends upon engagement of the T-cell receptor (TCR) and co-stimulatory signals provided by accessory cells. See, e.g., Jenkins and Johnson (1993) Curr. Opin. Immunol. 5:361-367; Bierer and Hahn (1993) Semin. Immunol. 5:249-261; June, et al. (1990) Immunol. Today 11:211-216; and Jenkins (1994) Immunity 1:443-446. A major, and well-studied, co-stimulatory interaction for T cells involves either CD28 or CTLA-4 on T cells with either B7 or B70 (Jenkins (1994) Immunity 1:443-446). Recent studies on CD28 deficient mice (Shahinian, et al. (1993) Science 261:609-612; Green, et al. (1994) Immunity 1:501-508) and CTLA-4 immunoglobulin expressing transgenic mice (Ronchese, et al. (1994) J. Exp. Med. 179:809-817) have revealed deficiencies in some T-cell responses though these mice have normal primary immune responses and normal CTL responses to lymphocytic choriomeningitis virus and vesicular stomatitis virus. As a result, both these studies conclude that other co-stimulatory molecules must be supporting T-cell function. However, identification of these molecules which mediate distinct costimulatory signals has been difficult.

Tumor Necrosis Factor (TNF) is the prototypic member of an emerging family of cytokines that function as prominent mediators of immune regulation and the inflammatory response. These ligands are typically type II membrane proteins, with homology at the carboxy terminus. A proteolytic processed soluble protein often is produced. See, e.g., Smith, et al. (1994) Cell 76:959-962; Armitage (1994) Current Opinion in Immunology 6:407-413; Gruss and Dower (1995) Blood 85:3378-3404; Wiley, et al. (1995) Immunity 3:673-682; and Baker and Reddy (1996) Oncogene 12:1-9. Crucial roles for these family members are evidenced by a number of studies, and they are implicated in regulation of apoptosis, peripheral tolerance, Ig maturation and isotype switching, and

general B cell and T cell functions. See, e.g., Thomson (ed. 1994) The Cytokine Handbook Academic Press, San Diego, CA; Naismith and Sprang (1998) Trends Biochem. Sci. 23:74-79; Lucas, et al. (1997) J. Leukoc. Biol. 61:551-558; Reddi (1997) Cell 5 89:159-161; Van Deventer (1997) Gut 40:443-448; Jablonska (1997) Postepy. Hig. Med. Dosw. 51:567-575; Hill and Lunec (1996) Mol. Aspects Med. 17:455-509; Aderka (1996) Cytokine Growth Factor Rev. 7:231-240; Lotz, et al. (1996) J. Leukoc. Biol. 60:1-7; and Gruss and Dower (1995) Cytokines Mol. Ther. 1:75-105. These imply
10 fundamental roles in immune and developmental networks relevant to human therapeutic needs. The identification of ligands and cell surface receptors allow determination of pairs, which will be useful in modulating such signal transduction.

The discovery of new cell markers is always potentially
15 useful. Moreover, the inability to modulate activation signals prevents control of inappropriate developmental or physiological responses in the immune system. The present invention provides at least one alternative costimulatory molecule, which will be useful as a marker for cell types, and agonists and antagonists of which
20 will be useful in modulating a plethora of immune conditions or responses.

The circulating component of the mammalian circulatory system comprises various cell types, including red and white blood cells of the erythroid and myeloid cell lineages. See, e.g., Rapaport
25 (1987) Introduction to Hematology (2d ed.) Lippincott, Philadelphia, PA; Jandl (1987) Blood: Textbook of Hematology, Little, Brown and Co., Boston, MA.; and Paul (ed. 1993) Fundamental Immunology (3d ed.) Raven Press, N.Y.

For some time, it has been known that the mammalian immune
30 response is based on a series of complex cellular interactions, called the "immune network." Recent research has provided new insights into the inner workings of this network. While it remains clear that much of the response does, in fact, revolve around the network-like interactions of lymphocytes, macrophages,
35 granulocytes, and other cells, immunologists now generally hold the opinion that soluble proteins, known as lymphokines,

cytokines, or monokines, play a critical role in controlling these cellular interactions. Thus, there is considerable interest in the isolation, characterization, and mechanisms of action of cell modulatory factors, an understanding of which should lead to significant advancements in the diagnosis and therapy of numerous medical abnormalities, e.g., immune system and other disorders.

Lymphokines apparently mediate cellular activities in a variety of ways. They have been shown to support the proliferation, growth, and differentiation of the pluripotential hematopoietic stem cells into vast numbers of progenitors comprising diverse cellular lineages making up a complex immune system. These interactions between the cellular components are necessary for a healthy immune response. These different cellular lineages often respond in a different manner when lymphokines are administered in conjunction with other agents.

The chemokines are a large and diverse superfamily of proteins. The superfamily is subdivided into two classical branches, based upon whether the first two cysteines in the chemokine motif are adjacent (termed the "C-C" branch), or spaced by an intervening residue ("C-X-C"). A more recently identified branch of chemokines lacks two cysteines in the corresponding motif, and is represented by the chemokines known as lymphotactins. Another recently identified branch has three intervening residues between the two cysteines, e.g., CX3C chemokines. See, e.g., Schall and Bacon (1994) Current Opinion in Immunology 6:865-873; and Bacon and Schall (1996) Int. Arch. Allergy & Immunol. 109:97-109.

Because the physiology mediated by these soluble molecules is so important, the discovery of novel chemokines will be important, both in diagnostic and therapeutic contexts.

In addition, while the general importance of the regulation of protein synthesis is universally accepted, the general importance of protein degradation has not been fully appreciated. One mechanism of protein degradation is via ubiquitination signals and degradation pathways. Ubiquitin (Ub) is a highly conserved 76 amino acid polypeptide that plays an important role in the regulation of protein degradation, cell-cycle progression, gene

transcription and signal transduction. The ubiquitination pathway is fine tuned and controlled, in part, by deubiquitination enzymes, which remove ubiquitin from proteins. Misregulation of the ubiquitination pathway may contribute problems in the protein quantity regulation, which may be associated, e.g., with malignant transformation, and oncogenesis through oncogenic counterparts of normally processed ubiquitinated proteins. Other clinical problems will often result from excessive or insufficient protein levels. Therefore, understanding the ubiquitination roles, e.g., in immune function, will increase our understanding of cell biology, which should have relevance, e.g., to malignant transformation.

Furthermore, growth of normal resting B cells (also referred to as "B lymphocytes") involves two distinct steps. First, the resting cells are activated to pass from the G₀ to G₁ phase of the cell cycle. See, e.g., Alberts, et al. (eds. 1989) Molecular Biology of the Cell Garland Publ., NY; and Darnell, et al. (1990) Molecular Cell Biology Freeman, NY. Next, the activated cells are induced to proliferate. See, e.g., Paul, ed. (1989) Fundamental Immunology, 2nd ed., Raven Press, NY; and the third edition. Several factors have been identified that induce growth of B cells, including interleukin-1 (IL-1), IL-2, IL-4, IL-10, and IL-13. In addition, antibodies against certain B cell surface molecules have been demonstrated to promote B cell proliferation. T cells (also referred to as "T lymphocytes") are also induced to proliferate by certain factors, which include phytohemagglutinin, anti-T cell receptor monoclonal antibodies, anti-CD3 monoclonal antibodies, and other agents.

B7 (CD80) and B70 (CD86) are the second "group" of molecules which strongly mediate B and T cell interaction. These molecules, on B cells, interact with their ligands CD28 and CTLA-4 on T cells. These interactions are major co-stimulatory signals for activation of both B and T cells.

During the last 15 years, it has become apparent that B7 (CD80) and B70 (CD86) play fundamental functions in T cell and B cell activation. Numerous in vitro and in vivo experiments have demonstrated that these two pairs of molecules represent important

targets for immunosuppression. See, e.g., Banchereau, et al. (1994) Ann. Rev. Immunol. 12:881-922; van Kooten, et al. (1996) Adv. Immunol. 61:1-77; Linsley and Ledbetter (1993) Ann. Rev. Immunol. 11:191-212).

5 In 1995, another molecule called RP105 was cloned from mouse splenic cells. See Miyake, et al (1995) J. Immunol. 154:3333-3340. Monoclonal antibodies against RP105 also induce strong proliferation of mouse B cells and protects mouse B cells from irradiation-induced apoptosis in a similar fashion to anti-CD40
10 antibody or CD40-ligand. See Miyake, et al. (1994) J. Exp. Med. 180:1217-1224.

The RP105 molecule and its ligand MD-1 may be an additional pair of molecules that play key roles in the activation of T cells and B cells. See Miyake, et al. (1998) J. Immunol. 161:1349-1353;
15 and Chan, et al., (1998) J. Exp. Med. 188:93-101 However, the human sequence of MD-1, has remained undetermined. The present invention provides this and also provides a previously undescribed second human homolog of mouse MD-1, (i.e., MD-2).

Many factors have been identified which influence the
20 differentiation process of precursor cells, or regulate the physiology or migration properties of specific cell types. These observations indicate that other factors exist whose functions in immune function were heretofore unrecognized. These factors provide for biological activities whose spectra of effects may be
25 distinct from known differentiation or activation factors. The absence of knowledge about the structural, biological, and physiological properties of the regulatory factors which regulate cell physiology in vivo prevents the modulation of the effects of such factors. Thus, medical conditions where regulation of the
30 development or physiology of relevant cells is required remains unmanageable.

Thus, significant therapeutic needs exist in the areas of cytokine regulation of physiology, protein degradation, and B cell signaling. The present invention provides important insights and
35 developments in these areas.

Cancer can occur in many tissues of the body. It results from a change in certain cells that causes them to evade the

normal growth limiting mechanisms, e.g., to escape the feedback controls that normally stop cellular growth and reproduction after a given number of such cells have developed. Cell division and transcription are highly coordinated processes that play important roles in this feedback control. See, e.g., Beeson, et al. (eds. 5 Textbook of Medicine, 15th ed., W.B. Saunders Co., Philadelphia, PA.; DeVita, et al. (eds. 1997) Cancer: Principles and Practice of Oncology, 5th ed., Lippincott, Philadelphia, PA; Neal and Hoskin (1997) Clinical Oncology: Basic Principles and Practice Oxford University Press, NY; Kastan (1997) Checkpoint Controls and Cancer CSH Press, NY; and Thomas (ed. 1996) Apoptosis and Cell Cycle Control in Cancer: Basic Mechanisms and Implications for Treating Malignant Disease BIOS Scientific, Oxford.

15 Molecules which function to regulate cell division play important roles in the controlled growth of various types of cells. Aberrations in these controls can lead to various disease states, e.g., oncogenesis, improper wound healing, developmental abnormalities, and metabolic problems.

20 The cell cycle can be divided into four phases: the presynthetic phases (G₀ and G₁); the phase of DNA synthesis (S); and the postsynthetic phase (G₂). See, e.g., Guyton (ed. 1976) Textbook of Medical Physiology, 5th ed., W.B. Saunders Co., Philadelphia, PA.; Alberts, et al. (eds. 1994) Molecular Biology of the Cell, 3rd ed., Garland Publishing, New York, NY; and 25 Darnell, et al. (eds. 1990) Molecular Cell Biology, 2nd ed., W.H. Freeman, New York, NY. Effective chemotherapeutic agents are often those which target diseased cells in the S phase, e.g., choriocarcinoma, acute lymphocytic leukemia, lymphocytic 30 lymphosarcoma, Burkitt's lymphoma, Hodgkin's disease, testicular neoplasms, Wilm's tumor, and Ewing's sarcoma. Unfortunately, oncogenic cells not actively dividing are less sensitive to these agents.

The lack of knowledge regarding the control of the cell cycle 35 has hampered the ability of medical science to specifically regulate cell division or immune responses. The present invention

provides compositions which will be important in the control of cell division and transcription.

SUMMARY OF THE INVENTION

5

The present invention is based, in part, upon the characterization of the genes and products relating to the DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5 chemokine, Dub11, Dub12, MD-1, MD-2, and cyclin E2. It provides nucleic acids, polypeptides, antibodies, and methods for making and using such compositions.

10

In the DC-PGT embodiments, the invention provides an isolated or recombinant antigenic polypeptide comprising: a plurality of distinct segments, wherein each segment has identity to at least 12 contiguous amino acids from the mature SEQ ID NO: 2; or at least 17 contiguous amino acids from the mature SEQ ID NO: 2. In certain embodiments, the plurality of segments includes one of at least 19 contiguous amino acids; or two of at least 15 contiguous amino acids. Other polypeptides include those wherein the polypeptide: comprises the mature SEQ ID NO: 2; binds with specificity to a polyclonal antibody which specifically binds to SEQ ID NO: 2; or the polypeptide: is a natural allelic variant of SEQ ID NO: 2; is at least 30 amino acids in length; exhibits at least two non-overlapping epitopes specific for SEQ ID NO: 2; is a synthetic polypeptide; is attached to a solid substrate; or is a 5-fold or less conservative substitution from SEQ ID NO: 2. Fusion polypeptides are also provided, e.g., comprising first and second portions, the first portion comprising a sequence as described and the second portion comprising a detectable marker. Pharmaceutical compositions are made available, e.g., comprising a sterile polypeptide, as described, in a pharmaceutically acceptable carrier.

20

25

Polynucleotide embodiments include an isolated or recombinant polynucleotide encoding a described polypeptide. Preferred forms will be such a polynucleotide which: comprises the mature polypeptide coding portion of SEQ ID NO: 1; or encodes the mature SEQ ID NO: 2. Preferred embodiments include wherein the polynucleotide is: a PCR product; a hybridization probe; a

30

35

mutagenesis primer; or made by chemical synthesis. Alternatively, the polynucleotide is: detectably labeled; a deoxyribonucleic acid; or double stranded. Also provided is an expression vector: comprising the described polynucleotide, including wherein the polypeptide specifically binds polyclonal antibodies generated against an immunogen of mature SEQ ID NO: 2; which selectively hybridizes under stringent hybridization conditions to a target polynucleotide sequence having at least 60 contiguous nucleotides from SEQ ID NO: 1; encodes a polypeptide having at least 50 contiguous amino acid residues from mature SEQ ID NO: 2; or is suitable for transfection into a prokaryote or eukaryote host cell. Preferably, the host cell is: a mammalian cell; a bacterial cell; an insect cell; a prokaryote; a eukaryote; or a COS cell. A method is provided, e.g., of making a polypeptide comprising expressing the vector in the host cell.

Other polynucleotides include an isolated or recombinant polynucleotide which hybridizes to the coding portion of SEQ ID NO: 1 under stringent hybridization and wash conditions of at least 50' C, a salt concentration of less than 400 mM, and 50% formamide. Such a nucleic acid may be an expression vector, which may hybridize to the coding portion of SEQ ID NO: 1 under stringent hybridization and wash conditions of at least 60' C, a salt concentration of less than 200 mM, and 50% formamide. Preferably, the vector encodes a polypeptide which specifically binds an antibody generated against a mature SEQ ID NO: 2. Another embodiment will be such a polynucleotide which hybridizes to SEQ ID NO: 1, wherein the polynucleotide is: a PCR product; a hybridization probe; a mutagenesis primer; or made by chemical synthesis.

Methods are provided, e.g., of modulating the physiology or development of a cell, comprising contacting the cell with an agonist or antagonist of a described polypeptide; of detecting the presence of a complementary polynucleotide in a sample, comprising contacting a described polynucleotide that selectively hybridizes with the complementary polynucleotide in the sample to form a detectable duplex; thereby indicating the presence of the polynucleotide in the sample; or for identifying a compound that

binds to a described polypeptide, comprising: incubating components comprising the compound and the polypeptide under conditions sufficient to allow the components to interact; and measuring the binding of the compound to the polypeptide.

5 In TNF receptor-like embodiments, the invention further provides an isolated or recombinant polynucleotide encoding an antigenic polypeptide comprising at least 17 contiguous amino acids from: the mature polypeptide from SEQ ID NO: 6; the mature polypeptide from SEQ ID NO: 8; the mature polypeptide from SEQ ID
10 NO: 10; the mature polypeptide from SEQ ID NO: 12; the mature polypeptide from SEQ ID NO: 17; the mature polypeptide from SEQ ID NO: 19; the mature polypeptide from SEQ ID NO: 21; or the mature polypeptide from SEQ ID NO: 23. In preferred embodiments, such polynucleotide will encode all of the polypeptide of: signal
15 processed SEQ ID NO: 6; signal processed SEQ ID NO: 8; signal processed SEQ ID NO: 10; signal processed SEQ ID NO: 12; signal processed SEQ ID NO: 17; SEQ ID NO: 19; SEQ ID NO: 21; or SEQ ID NO: 23. Other embodiments include such a polynucleotide, which hybridizes at 55° C, less than 500 mM salt, and 50% formamide to
20 the: mature protein coding portion of SEQ ID NO: 5; signal processed coding portion of SEQ ID NO: 7; signal processed coding portion of SEQ ID NO: 9; signal processed coding portion of SEQ ID NO: 11; mature protein coding portion of SEQ ID NO: 16; polypeptide coding portion of SEQ ID NO: 18; polypeptide coding
25 portion of SEQ ID NO: 20; or polypeptide coding portion of SEQ ID NO: 22. Other forms include those polynucleotides, comprising at least 35 contiguous nucleotides of: mature protein coding portion of SEQ ID NO: 5; signal processed coding portion of SEQ ID NO: 7; signal processed coding portion of SEQ ID NO: 9; signal processed
30 coding portion of SEQ ID NO: 11; mature protein coding portion of SEQ ID NO: 16; polypeptide coding portion of SEQ ID NO: 18; polypeptide coding portion of SEQ ID NO: 20; or polypeptide coding portion of SEQ ID NO: 22. Various expression vectors are provided comprising such a polynucleotide. The invention also provides a
35 host cell containing the expression vector, including a eukaryotic cell.

Methods are provided, e.g., making an antigenic polypeptide comprising expressing a recombinant polynucleotide; for detecting a polynucleotide, comprising contacting the polynucleotide with a probe that hybridizes, under stringent conditions, to at least 25
5 contiguous nucleotides of the: mature protein coding portion of SEQ ID NO: 5; signal processed coding portion of SEQ ID NO: 7; signal processed coding portion of SEQ ID NO: 9; signal processed coding portion of SEQ ID NO: 11; mature protein coding portion of SEQ ID NO: 16; polypeptide coding portion of SEQ ID NO: 18;
10 polypeptide coding portion of SEQ ID NO: 20; or polypeptide coding portion of SEQ ID NO: 22; to form a duplex, wherein detection of the duplex indicates the presence of the polynucleotide. Kits are provided, e.g., for the detection of a described polynucleotide, comprising a compartment containing a probe that hybridizes, under
15 stringent hybridization conditions, to at least 17 contiguous nucleotides of a described polynucleotide to form a duplex. Preferably, the probe is detectably labeled.

Binding compounds are provided, including antibodies, comprising an antibody binding site which specifically binds to a
20 polypeptide comprising at least 17 contiguous amino acids from: signal processed SEQ ID NO: 6; signal processed SEQ ID NO: 8; signal processed SEQ ID NO: 10; signal processed SEQ ID NO: 12; signal processed SEQ ID NO: 17; SEQ ID NO: 19; SEQ ID NO: 21; or SEQ ID NO: 23. Preferably, the antibody binding site is:
25 selectively immunoreactive with the: signal processed SEQ ID NO: 6; signal processed SEQ ID NO: 8; signal processed SEQ ID NO: 10; signal processed SEQ ID NO: 12; signal processed SEQ ID NO: 17; SEQ ID NO: 19; SEQ ID NO: 21; or SEQ ID NO: 23; raised against a purified or recombinantly produced human HDTEA84 protein; raised
30 against a purified or recombinantly produced human HSLJD37R protein; or in a monoclonal antibody, Fab, or F(ab)₂; or the binding compound is: an antibody molecule; a polyclonal antiserum; detectably labeled; sterile; or in a buffered composition.

Such compositions allow various methods, including using the
35 binding compound, comprising contacting the binding compound with a biological sample comprising an antigen, thereby forming a binding compound:antigen complex. Preferably, the biological

sample is from a human, and the binding compound is an antibody. Such also allow for production of a detection kit comprising the binding compound, and: instructional material for the use of the binding compound for the detection; or a compartment providing
5 segregation of the binding compound.

Polypeptides are also made available, e.g., a substantially pure or isolated antigenic polypeptide, which binds to the described binding composition, and further comprises at least 17 contiguous amino acids from: signal processed SEQ ID NO: 6; signal
10 processed SEQ ID NO: 8; signal processed SEQ ID NO: 10; signal processed SEQ ID NO: 12; signal processed SEQ ID NO: 17; SEQ ID NO: 19; SEQ ID NO: 21; or SEQ ID NO: 23. Preferred polypeptides include those which: comprise at least a fragment of at least 25 contiguous amino acid residues from: a primate HDTEA84 protein; a
15 primate HSLJD37R protein; or a rodent or primate RANKL protein; or are soluble polypeptides; are detectably labeled; are in a sterile composition; are in a buffered composition; bind to an sialic acid residue; are recombinantly produced; or have a naturally occurring polypeptide sequence. In other embodiments, the polypeptide
20 comprises at least 17 contiguous amino acids from the: signal processed SEQ ID NO: 6; signal processed SEQ ID NO: 8; signal processed SEQ ID NO: 12; signal processed SEQ ID NO: 17; SEQ ID NO: 19; SEQ ID NO: 21; or SEQ ID NO: 23.

Methods are provided, including a method of modulating a precursor cell physiology or function comprising a step of
25 contacting the cell with: a binding compound which binds to a described polypeptide; an HDTEA84 polypeptide; an HSLJD37R polypeptide; or a RANKL polypeptide. The method may be one wherein the contacting is in combination with a TNF family ligand,
30 or an antagonist of the TNF family ligand.

In other embodiments, the present invention provides compositions related to other chemokine, Dub, or surface protein genes. Polypeptide embodiments include: a substantially pure or recombinant HCC5 polypeptide exhibiting identity over a length of
35 at least 12 amino acids to SEQ ID NO: 25; an isolated natural sequence HCC5 of mature SEQ ID NO: 25; a fusion protein comprising HCC5 sequence; a substantially pure or recombinant Dub11

polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 32 or 34; an isolated natural sequence Dub11 of mature SEQ ID NO: 32 or 34; a fusion protein comprising Dub11 sequence; a substantially pure or recombinant Dub12

5 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 36 or 38; an isolated natural sequence Dub12 of mature SEQ ID NO: 36 or 38; a fusion protein comprising Dub12 sequence; a substantially pure or recombinant MD-1

10 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 42; an isolated natural sequence MD-1 of mature SEQ ID NO: 42; a fusion protein comprising primate MD-1 sequence; a substantially pure or recombinant MD-2 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 44 or 46; an isolated natural sequence MD-2 of

15 mature SEQ ID NO: 44 or 46; a fusion protein comprising primate MD-2 sequence; a substantially pure or recombinant MD-2 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 48 or 49; an isolated natural sequence MD-2 of mature SEQ ID NO: 48; or a fusion protein comprising

20 murine MD-2 sequence. Preferred embodiments include substantially pure or isolated polypeptides which match the sequences over a stretch of at least 17 amino acids; more preferably over a stretch of at least 21 amino acids; over 25, 30, 35, 50, 75 or more. In other preferred embodiments, the HCC5 polypeptide: is from a

25 primate, including a human; comprises at least one polypeptide segment of SEQ ID NO: 25; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of HCC5; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a primate HCC5;

30 exhibits a sequence identity over a length of at least 35 amino acids to a HCC5; is glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence; or the

35 Dub11 polypeptide: is from a primate, including a human; comprises at least one polypeptide segment of SEQ ID NO: 32 or 34; exhibits a plurality of portions exhibiting the identity; is a natural

allelic variant of Dub11; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a primate Dub11; exhibits a sequence identity over a length of at least about 35 amino acids to a Dub11; is

5 glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence; or the Dub12 polypeptide: is from a primate, including a human; comprises at

10 least one polypeptide segment of SEQ ID NO: 36 or 38; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of Dub12; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a primate Dub12; exhibits a sequence identity over a

15 length of at least about 35 amino acids to a Dub12; is glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence; or the primate MD-1

20 polypeptide: is from a human; comprises at least one polypeptide segment of SEQ ID NO: 42; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of primate MD-1; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a

25 primate MD-1; exhibits a sequence identity over a length of at least about 35 amino acids to a primate MD-1; is glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion

30 variant from a natural sequence; or the primate MD-2 polypeptide: is from a human; comprises at least one polypeptide segment of SEQ ID NO: 44 or 46; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of primate MD-2; has a length at least about 30 amino acids; exhibits at least two non-

35 overlapping epitopes which are specific for a primate MD-2; exhibits a sequence identity over a length of at least about 35 amino acids to a primate MD-2; is glycosylated; is a synthetic

polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence; or the rodent MD-2 polypeptide: is from a mouse; comprises at least one polypeptide segment of SEQ ID NO: 48 or 49; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of rodent MD-2; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a rodent MD-2; exhibits a sequence identity over a length of at least about 35 amino acids to a rodent MD-2; is glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence.

Sterile compositions comprising such polypeptides are also provided, along with those comprising: the HCC5 polypeptide and: a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; another chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4; or an antibody antagonist for a chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4; the Dub11 polypeptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; the Dub12 polypeptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; the MD-1 polypeptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; the MD-2 polypeptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

Fusion proteins are provided, e.g., comprising: mature protein sequence of SEQ ID NO: 25; mature protein sequence of SEQ

ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36 or SEQ ID NO: 38; mature protein sequence of SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, or SEQ ID NO: 49; a detection or purification tag, including a FLAG, His6, or Ig sequence; or sequence of another chemokine protein with the chemokine polypeptide Kits are provided, e.g., comprising a described polypeptide and: a compartment comprising the polypeptide; and/or instructions for use or disposal of reagents in the kit.

Binding compounds, including antibodies, are provided, e.g., comprising an antigen binding portion from an antibody, which specifically binds to a natural: HCC5 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature HCC5 polypeptide sequence of SEQ ID NO: 25; is raised against a mature HCC5; is raised to a purified HCC5; is immunoselected; is a polyclonal antibody; binds to a denatured HCC5; or exhibits a Kd to HCC5 antigen of at least 30 μ M; or Dub11 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature Dub11 polypeptide sequence of SEQ ID NO: 32 or SEQ ID NO: 34; is raised against a mature Dub11; is raised to a purified Dub11; is immunoselected; is a polyclonal antibody; binds to a denatured Dub11; or exhibits a Kd to Dub11 antigen of at least 30 μ M; or Dub12 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature Dub12 polypeptide sequence of SEQ ID NO: 36 or SEQ ID NO: 38; is raised against a mature Dub12; is raised to a purified Dub12; is immunoselected; is a polyclonal antibody; binds to a denatured Dub12; or exhibits a Kd to Dub12 antigen of at least 30 μ M; or a primate MD-1 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature polypeptide sequence of SEQ ID NO: 42; is raised against a mature MD-1; is raised to a purified MD-1; is immunoselected; is a polyclonal antibody; binds to a denatured MD-1; or exhibits a Kd to MD-1 antigen of at least 30 μ M; or a primate MD-2 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature MD-2 polypeptide sequence of SEQ ID NO: 44, or SEQ ID NO: 46; is raised against a mature MD-2; is raised to a purified MD-2; is immunoselected; is a polyclonal antibody; binds to a denatured

MD-2; or exhibits a K_d to MD-2 antigen of at least 30 μM ; or a rodent MD-2 polypeptide, wherein the antibody: is raised against a peptide sequence of a mature MD-2 polypeptide sequence of SEQ ID NO: 48, or SEQ ID NO: 49; is raised against a mature rodent MD-2; is raised to a purified rodent MD-2; is immunoselected; is a polyclonal antibody; binds to a denatured rodent MD-2; or exhibits a K_d to antigen of at least 30 μM . In certain embodiments, the binding composition will be one wherein: the polypeptide is from a primate or rodent; the binding compound is an Fv, Fab, or Fab2 fragment; the binding compound is conjugated to another chemical moiety; is attached to a solid substrate, including a bead or plastic membrane; is in a sterile composition; or is detectably labeled, including a radioactive or fluorescent label.

Kits are provided comprising the binding compound, and: a compartment comprising the binding compound; a compartment comprising purified antigen; and/or instructions for use or disposal of reagents in the kit. Methods are provided for producing an antigen:antibody complex, comprising contacting an antibody and: a primate HCC5 polypeptide; a primate Dub11 polypeptide; a primate Dub12 polypeptide; a primate MD-1 polypeptide; a primate MD-2 polypeptide; or a rodent MD-2 polypeptide; thereby allowing the complex to form. Other compositions are provided, e.g., the binding compound and: a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; or an antibody antagonist for another chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4.

Nucleic acid embodiments include, e.g., an isolated or recombinant nucleic acid encoding a polypeptide or fusion protein described, wherein: the HCC5: polypeptide is from a primate, including a human; or nucleic acid: encodes an antigenic HCC5 peptide sequence of SEQ ID NO: 25 encodes a plurality of antigenic HCC5 peptide sequences of SEQ ID NO: 25; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the HCC5 segment; is a hybridization probe for a gene encoding the HCC5 polypeptide;

or further encodes another chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4; or the Dub11: polypeptide is from a primate, including a human; or nucleic acid: encodes a Dub11 antigenic peptide sequence of SEQ ID NO: 32; or SEQ ID NO: 34; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 32 or SEQ ID NO: 34; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the Dub11 segment; or is a hybridization probe for a gene encoding the Dub11 polypeptide; the Dub12: polypeptide is from a primate, including a human; or nucleic acid: encodes an antigenic Dub12 peptide sequence of SEQ ID NO: 36 or SEQ ID NO: 38; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 36 or SEQ ID NO: 38; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the DUB12 segment; is a hybridization probe for a gene encoding the Dub12 polypeptide; or the primate MD-1: polypeptide is from a primate, including a human; or nucleic acid: encodes an antigenic MD-1 peptide sequence of SEQ ID NO: 42; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 42; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the MD-1 segment; is a hybridization probe for a gene encoding the Dub11 polypeptide; or the primate MD-2: polypeptide is from a human; or nucleic acid: encodes an antigenic MD-2 peptide sequence of SEQ ID NO: 44, or SEQ ID NO: 46; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 44, or SEQ ID NO: 46; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the segment; is a hybridization probe for a gene encoding the primate MD-2 polypeptide; or the rodent MD-2: polypeptide is from a mouse; or nucleic acid: encodes an antigenic MD-2 peptide sequence of SEQ ID NO: 48, or SEQ ID NO: 49; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 48, or SEQ ID NO: 49; exhibits identity over at least 25 nucleotides to a natural cDNA encoding the MD-2 segment; or is a hybridization probe for a gene encoding the rodent MD-2 polypeptide. Other nucleic acid embodiments include the described, which: is an expression vector; further comprises an origin of replication; is from a natural source; comprises a detectable label; comprises synthetic nucleotide sequence is less than 6 kb, preferably less than 3 kb; is from a

primate, including a human; comprises a natural full length coding sequence; or is a PCR primer, PCR product, or mutagenesis primer.

Various cells are provided, including a cell or tissue comprising a described recombinant nucleic acid, including wherein
5 the cell is: a prokaryotic cell; a eukaryotic cell; a bacterial cell; a yeast cell; an insect cell; a mammalian cell; a mouse cell; a primate cell; or a human cell.

Kits are provided, e.g., comprising a described nucleic acid, and: a compartment comprising the nucleic acid; a compartment
10 comprising a nucleic acid encoding another chemokine, including HCC1, HCC2, HCC3, and HCC4; or instructions for use or disposal of reagents in the kit.

Alternative nucleic acids include those which: hybridize under wash conditions of 45° C and less than 2M salt to the
15 polypeptide coding portion of SEQ ID NO: 24; hybridize under wash conditions of 45° C and less than 2M salt to the polypeptide coding portions of SEQ ID NO: 31 or 33; hybridize under wash conditions of 45° C and less than 2M salt to the coding portions of SEQ ID NO: 35 or 37; hybridize under wash conditions of 45° C
20 and less than 2M salt to the coding portion of SEQ ID NO: 41; hybridize under wash conditions of 45° C and less than 2M salt to the coding portion of SEQ ID NO: 43 or 45. or hybridize under wash conditions of 45° C and less than 2M salt to the coding portion of SEQ ID NO: 47. Preferably, the wash conditions are at 55° C
25 and/or 500 mM salt; or at 65° C and/or 150 mM salt.

Additionally, methods are provided, e.g., of modulating physiology or development of a cell or tissue culture cells comprising exposing the cell to an agonist or antagonist of HCC5,
primate MD-1, primate MD-2, or rodent MD-2. Others include
30 methods of detecting specific binding to a compound, comprising: contacting the compound to a composition selected from the group of: an antigen binding site which specifically binds to: an HCC5 chemokine; a Dub11; a Dub12; a primate MD-1; a primate MD-2; a rodent MD-2; or an expression vector encoding: an HCC5 chemokine
35 or fragment thereof; a Dub11 or fragment thereof; a Dub12 or fragment thereof; a primate MD-1 or fragment thereof; a primate MD-2 or fragment thereof; or a rodent MD-2 or fragment thereof; a

substantially pure protein which is specifically recognized by the antigen binding site of the described antigen binding sites; a substantially pure HCC5 chemokine or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of HCC5 chemokine sequence; a substantially pure Dub11 or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of Dub11 sequence; a substantially pure Dub12 or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of Dub11 sequence; a substantially pure primate MD-1 or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of primate MD-1 sequence; a substantially pure primate MD-2 or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of primate MD-2 sequence; a substantially pure rodent MD-2 or peptide thereof, or a fusion protein comprising a 30 amino acid sequence portion of rodent MD-2 sequence; and then detecting binding of the compound to the composition.

Particular polynucleotide embodiments include an isolated or recombinant polynucleotide which: encodes at least 17 contiguous amino acid residues of SEQ ID NO: 54; encodes at least two distinct segments of at least 10 contiguous amino acid residues of SEQ ID NO 54; or comprises one or more segments at least 21 contiguous nucleotides of SEQ ID NO: 53. Such polynucleotides allow methods of making: a polypeptide comprising expressing a described expression vector, thereby producing the polypeptide; a duplex nucleic acid comprising contacting a polynucleotide with a complementary nucleic acid, thereby resulting in production of the duplex nucleic acid; a synthetic polynucleotide, comprising chemically polymerizing nucleotides to produce the polynucleotide; or a polynucleotide comprising using a PCR method.

Cyclin polypeptide embodiments include an isolated or recombinant antigenic polypeptide comprising at least: one segment comprising at least 17 contiguous amino acids from SEQ ID NO: 54; or at least two distinct segments of at least 11 contiguous amino acids from SEQ ID NO: 54. Such polypeptide may: comprise at least one segment comprising at least 17 contiguous amino acids from SEQ ID NO: 54; and exhibit at least two non-overlapping epitopes which

are selective for primate protein of SEQ ID NO: 54. Other embodiments include those wherein the polypeptide: is a 5-fold or less substitution from a natural sequence; is a deletion or insertion variant from a natural sequence; or comprises at least two distinct segments of at least 11 contiguous amino acids from SEQ ID NO: 54. Preferably the polypeptide is antigenic, and will typically comprise at least one sequence from (SEQ ID NO: 54) KESRYVHD (residues 120-127), DKHFEVLH (residues 127-134), HSDLEPQM (residues 134-141), QKDINKNM (residues 177-184), YAPKLQEF (residues 203-210), SEEDILRM (residues 219-226), LRMEIIL (residues 224-231), ELCPVTII (residues 237-244), and LFLQVDAL (residues 249-256); and/or the segments of at least 11 contiguous amino acids comprise one the segment with at least 14 contiguous amino acids from SEQ ID NO: 54. Such polypeptides may further exhibit at least two non-overlapping epitopes which are selective for primate protein of SEQ ID NO: 54; and/or may: comprise a mature sequence of SEQ ID NO: 2; bind with selectivity to an antibody generated against an immunogen of SEQ ID NO: 54; comprise a plurality of polypeptide segments of 17 contiguous amino acids of SEQ ID NO: 54; or be a natural allelic variant of SEQ ID NO: 54. The polypeptide may: be in a sterile composition; have a length at least 30 amino acids; be not glycosylated; be denatured; be a synthetic polypeptide; be attached to a solid substrate; or be a fusion protein with a detection or purification tag, including a FLAG, His6, or Ig sequence. Other embodiments include those wherein the polypeptide: is a 5-fold or less substitution from a natural sequence; or is a deletion or insertion variant from a natural sequence.

Various kits are provided, e.g., which comprise such polypeptides and instructions for the use or disposal of the polypeptide or other reagents of the kit.

Methods are provided, e.g., to label the polypeptide, comprising labeling the polypeptide with a radioactive label; to separate the polypeptide from another polypeptide in a mixture, comprising running the mixture on a chromatography matrix, thereby separating the polypeptides; to identify a compound that binds selectively to the polypeptide, comprising incubating the compound

with the polypeptide under appropriate conditions; thereby causing the component to bind to the polypeptide; to conjugate the polypeptide to a matrix, comprising derivatizing the polypeptide with a reactive reagent, and conjugating the polypeptide to the matrix; or inducing an antibody response to the polypeptide, comprising introducing the polypeptide as an antigen to an immune system, thereby inducing the response.

Binding compounds are provided, e.g., antibodies, comprising an antigen binding portion from an antibody which binds with selectivity to described polypeptides. Methods are made available for evaluating the selectivity of binding of a compound to cyclin E2, comprising contacting the compound to a recombinant cyclin E2 polypeptide and at least one other cyclin; and comparing binding of the compound to the cyclins.

15

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

All references cited herein are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

I. General

It is to be understood that this invention is not limited to the particular compositions, methods, and techniques described herein, as such compositions, methods, and techniques may vary. It is to be understood that the terminology used herein is for the purpose of describing particular embodiments, and is not intended to limit the scope of the present invention which is to be limited by the appended claims.

As used herein, including the appended claims, singular forms of words such as "a," "an," and "the" include their corresponding plural referents unless the context clearly dictates otherwise. Thus, e.g., reference to "a polynucleotide" includes one or more different polynucleotides, reference to "a composition" includes one or more of such compositions, and reference to "a method" includes reference to equivalent steps and methods known to a person of ordinary skill in the art, and so forth.

Unless otherwise defined, technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references discussed above are provided for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate any such disclosure by virtue of its prior invention.

The present invention also provides amino acid sequences and DNA sequences encoding various mammalian proteins, e.g., which are polypeptides produced by selected cells. Among these proteins are

those which: mediate uptake of substrates, e.g., prostaglandin-like molecules, modulate or mediate, e.g., induce or prevent trafficking, proliferation, or differentiation of, interacting cells, or are intracellular proteins which are important in various cellular processes, e.g., deubiquitination of proteins or cell cycle regulation.

The Prostaglandin-like Transporter (PGT) of the present invention is expressed particularly in antigen presenting cells of the immune system, e.g., dendritic cells. As such, the transporter is designated a dendritic cell prostaglandin-like transporter (DC-PGT). Consequently, the DC-PGT of the present invention offers the means to establish fundamental understanding on the role of PG influence on immune function.

The present invention provides DNA sequence encoding a mammalian protein that exhibits structural features characteristic of functionally significant proteins, particularly which serve as organic anion transporters. This family of organic anion transporters includes: the prostaglandin transporters of man (Lu, et al. (1996) J. Clin. Invest. 98:1142-1149) and rat; organic anion transporters in man and rat; brain digoxin transporters and Matrin F/G of rat (Kanai, et al. (1995) Science 268:866-869).

Transporters of this family typically are 12 transmembrane proteins of approximately 650 amino acids in length. Characteristic of this group of proteins is a cysteine rich region located in one of the extracellular loops, which resembles a zinc finger motif. It is not entirely certain whether these polypeptides mediate primarily the influx or efflux of prostaglandins and organic anions, and they may, under different circumstances produce influx or efflux depending, e.g., on the intracellular concentration of the organic anions concerned.

The DC PGT protein of the present invention is closest in homology to the prostaglandin transporters and it is probable that a prostaglandin is the major anion transported. The human gene embodiment described herein, isolated as designate DC-PGT or clone 240, contains an open reading frame encoding a presumptive protein of about 709 amino acids. This protein exhibits intracellular, transmembrane, and extracellular protein segments, revealing novel

aspects of organic anion transport that may be relevant during mammalian development, e.g., development of dendritic cells of the immune system.

The introduction of evolutionary information in the form of sequence homologs simplifies the structural analysis considerably for related molecules which share a common structural framework in spite of considerable sequence divergence, see, e.g., Chothia and Lesk (1986) EMBO J. 5:823-826. This concept can be effectively extended to the strong prediction of TM regions across an aligned protein family, whereas any single sequence may provide an uncertain topology. See Persson and Argos (1994) J. Mol. Biol. 237:182-192; and Rost, et al. (1995) Protein Sci. 4:521-533. For the DC PGT, a number of sequence homologs were first assembled by comparative matching to protein and translated nucleotide databases (Altschul, et al. (1994) Nature Genet. 6:119-129; Koonin, et al. (1994) EMBO J. 13:493-503). These relatives of DC-PGT include a ubiquitously expressed PGT from primate, e.g., human (GenBank: locus HSU70867, accession U70867), and a PGT from rodent, e.g., rat (prostaglandin transporter - rat, GenBank Acc. No. 1083766; Kanai, et al. (1995) Science 268:866-869). These sequences were subjected to parallel analyses by a suite of computer programs that have greatly improved on the initial Kyte and Doolittle (1982) hydropathic profile as a means of predicting the topology of integral membrane proteins. Four algorithms (ALOM, MTOP, MEMSAT and TopPredII) (Klein, et al. (1985) Biochim. Biophys. Acta 815:468-476; Hartmann, et al. (1989) Proc. Nat'l Acad. Sci. USA 86:5786-5790; Jones, et al. (1994) Biochem. Biophys. Res. Commun. 201:330-334; and Claros and von Heijne (1994) Comp. Applic. Biosci. 10:685-686) were used to individually predict TM extensions and orientations; these predictions were pooled and mapped onto the multiple sequence alignment produced by ClustalW and MACAW (Thompson, et al. (1994) Nucl. Acids Res. 22:4673-4680; and Schuler, et al. (1991) Proteins 9:180-190). Furthermore, these multiply aligned sequence files were used as input to PHD and TMAP (Rost, et al. (1995) Protein Sci. 4:521-533; Persson and Argos (1994) J. Mol. Biol. 237:182-192) for a familial prediction

of shared TM regions. Structural features that persisted in this two-step analysis are likely to be shared topological traits present in all members of this organic anion transporter family.

HDTEA84, HSLJD37R, and RANKL genes and proteins are also provided, which are related to the TNF signaling pathways. The antigens HDTEA84, HSLJD37R, and RANKL, and fragments, or antagonists will be useful in physiological modulation of cells expressing receptors for, e.g., ligands of the TNF family. Some of these antigens appear to lack a membrane spanning segment, suggesting that they are soluble forms of receptor. This suggests that the soluble proteins can serve as antagonists of the TNF-like ligands. In addition, it is likely that membrane spanning forms exist, which serve as signaling receptors mediating cellular response to the ligands.

The HDTEA84 gene has been detected in cDNA libraries derived from Hodgkin's lymphoma, endothelial cells, keratinocytes, prostrate, and cerebellum. It exhibits significant sequence similarity to the osteoprotegerin ligand receptor reported by Lacey, et al. (1998) Cell 93:165-176. The HDTEA84 will likely modulate proliferation or development by antagonizing its respective ligand. Membrane associated forms should exist, likely alternatively spliced transcription products.

The HSLJD37R exhibits like similarity to receptors for TNF. While the first embodiment is an incomplete sequence, the available portion currently lacks an identified transmembrane segment. Additional efforts provide a full length sequence, and an alternative splice variant.

The rodent 427152#4 Rank-like (RANKL) was detected in a rodent cDNA library panel probed with Mouse 427152#4 (204 bp). Positive signals were detected in CH12 (B cell line); rag-1 thymus; rag-1 heart; rag-1 brain (best signal); rag-1 testes; rag-1 liver; normal lung; rag-1 lung; asthmatic lung; tolerized and challenged lung; Nippo-infected lung; Nippo IL-4 K.O. lung; Nippo anti-IL-5 treated lung; influenza lung; guinea pig allergic lung; w.t. stomach; and w.t. colon on a 3 day exposure at -80° C with an intensifier screen. On a 2 week exposure at -80° C with screen, signals were also detected in the following libraries: Mell14+

naive; Mel14+ Th1; Mel14+ Th2; Th1 3 week B1/6; large B cell;
bEnd3 + TNF α + IL-10, guinea pig normal lung; and Rag Hh- colon.

The primate, e.g., human, Rank-like (RANKL) homologs of
rodent 427152#4 were detected in a human cDNA library panel probed
5 with mouse 427152#4 (204 bp). Signals were detected in monkey
asthma lung 4 h (1.6-2.0 kb) and adult placenta (2.5-3.0 kb) on a
3 day exposure at -80° C with screen. On a 2 week exposure at -
80° C with screen, signals were also detected in the following
libraries: CD1a+ 95% DC activated CHA (kidney epithelial carcinoma
10 cell line); monkey lung normal; psoriasis skin; fetal lung; fetal
ovary; fetal testes; and fetal spleen.

Each of these proteins will also be useful as antigens, e.g.,
immunogens, for raising antibodies to various epitopes on the
protein, linear and/or conformational epitopes. The molecules may
15 be useful in defining various cell subsets, either by the
molecules produced by, or by expression of membrane forms of the
receptors. Such cells should be responsive to the respective
ligands. Soluble forms of the receptors should serve as
antagonists of the ligand, binding to the ligand and preventing
20 interaction with membrane forms, which would mediate signaling.

Both genes express proteins which exhibit structural motifs
characteristic of a member of the TNF receptor family. SEQ ID NO: 5
and SEQ ID NO: 6, respectively, provide the nucleic acid and
predicted amino acid sequences for primate, e.g., human, HDTEA84.
25 SEQ ID NO: 7 and SEQ ID NO: 8, respectively, provide the nucleic
acid and predicted amino acid sequences for primate, e.g., human,
HSLJD37R.

Interesting features of the HDTEA84 include: signal sequence
from about 1-11; TNF receptor Cys rich domains I (about 32-72), II
30 (about 73-113), III (about 114-150), and IV (about 151-193); and
unique region from about 194-300. Features for the HSLJD37R (SEQ
ID NO: 10 form), partly based on alignment with HDTEA84: signal
sequence from about 1-41; TNF receptor Cys rich domains I (about
42-90), II (about 91-131), III (about 132-168), and IV (about 169-
35 211); transmembrane segment from about 354-370. Similar alignment
of the other variants will identify similar features. Segments
including combinations or excluding such segments may be desired.

The structural homology of HDTEA84, HSLJD37R, and RANKL to members of the TNF receptor family suggests related function of these molecules. See, e.g., Lacey, et al. (1998) Cell 93:165-176. The sequences, however, mostly lack a transmembrane segment, suggesting that the proteins are soluble receptor forms. They may well also have membrane bound forms resulting, e.g., from alternatively spliced transcript variants. The soluble forms are likely to be antagonists of the ligand, e.g., blocking the binding of ligand to a membrane bound form of signaling receptor. Thus, these molecules may be useful in the treatment of abnormal immune or developmental disorders.

The natural antigens should be capable of modulating various biochemical responses which lead to biological or physiological responses in target cells. The embodiments characterized herein are from primate, e.g., human, but other species variants almost surely exist, e.g., rodents, etc. See below. The descriptions below are directed, for exemplary purposes, to primate HDTEA84, HSLJD37R, or RANKL, but are likewise applicable to related embodiments from other species.

The HDTEA84, HSLJD37R, and RANKL clones were assembled through the careful analysis of ESTs present in various databases, e.g., Merck-WashU public database. These genes exhibit structural motifs characteristic of a member of the TNF receptor family. Compare, e.g., with the TNF receptor, NGF-receptor, and FAS receptor. Table 3 discloses the nucleic acid and predicted amino acid sequences for primate, e.g., human, HDTEA84. The ESTs were identified from several different libraries.

SEQ ID NO: 7 AND SEQ ID NO: 8, respectively, disclose partial nucleic acid and predicted amino acid sequences for primate, e.g., human, HSLJD37R. The ESTs were identified from several different libraries derived from: smooth muscle, pancreas tumor, adipocytes, HUVEC cells, adult pulmonary, endothelial cells, prostate cell line PC3, microvascular endothelial cells, fetal heart, and dendritic cells. Other sequences were detected in libraries from: multiple sclerosis lesions, breast, kidney, and germinal center B cells.

SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20 and SEQ ID NO: 22 provide the sequences of various mammalian genes designated RANKL.

Interesting features of the rodent RANKL include: signal sequence from about 1-29; TNF receptor Cys rich domain I (about 33-74), II (about 75-114), and III (about 115-135). Interesting features of the primate RANKL include: TNF receptor Cys rich domain I (about 1-43), II (about 44-83), and III (about 84-104); transmembrane segment from about 139-155. Alignment with other TNF receptors will identify additional interesting corresponding features. Segments with boundaries at these positions may be especially interesting.

Hybridization signals with RANKL were detected with rodent, e.g., mouse sequence, in CH12 (B cell line), rag-1 thymus, rag-1 heart, rag-1 brain (strongest signal), rag-1 testes, rag-1 liver, normal lung, rag-1 lung, asthmatic lung, tolerized and challenged lung, Nippo-infected lung, Nippo IL-4 K.O. lung, Nippo anti-IL-5 lung, influenza lung, guinea pig allergic lung, w.t. stomach, and w.t. colon on a 3 day exposure at -80° C with a screen. On a 2 week exposure at -80° C with screen, signals were also detected in the following libraries: Mel 14+ naive, Mel14+ Th1, Mel14+ Th2, Th1 3 week B1/6, large B cell, bEnd3 + TNF α + IL-10, guinea pig normal lung, and Rag Hh- colon. Probes of human libraries with rodent sequence provided: detectable signals in Monkey asthma lung 4 h (1.6-2.0 kb) and adult placenta (2.5-3.0 kb) on a 3 day exposure at -80° C with screen. On a 2 week exposure at -80° C with screen, signals were also detected in the following libraries: CD1a+ 95% DC activated, CHA (kidney epithelial carcinoma cell line), monkey lung normal, psoriasis skin, fetal lung, fetal ovary, fetal testes, and fetal spleen.

In another embodiment, the invention provides a chemokine. For a review of the chemokine family, see, e.g., Lodi, et al. (1994) Science 263:1762-1767; Gronenborn and Clore (1991) Protein Engineering 4:263-269; Miller and Kranger (1992) Proc. Nat'l Acad. Sci. USA 89:2950-2954; Matsushima and Oppenheim (1989) Cytokine 1:2-13; Stoeckle and Baker (1990) New Biol. 2:313-323; Oppenheim, et al. (1991) Ann. Rev. Immunol. 9:617-648; Schall (1991) Cytokine

3:165-183; and Thomson (ed. 1994) The Cytokine Handbook 2d ed. Academic Press, NY.

The new chemokine described herein is designated HCC5 which is a CC chemokine. See SEQ ID NO: 24 and SEQ ID NO: 25. The descriptions are directed, for exemplary purposes, to the human HCC5 natural allele described, but are likewise applicable to allelic and/or polymorphic variants, e.g., from other individuals, as well as splicing variants, e.g., natural forms. Based on sequence analysis of the chemokine protein sequences described below, it is apparent that HCC5 belongs to the CC chemokine family. See, e.g., stem cell mobilizing chemokine (CKbeta-1) from Kreider, et al. (1997) Patent WO 9715594 (SEQ ID NO: 26) and GenBank Accession number 97P-W17659; macrophage inflammatory protein-1-gamma (MIP-1) from Adams, et al. (1995) Patent WO 9517092 (SEQ ID NO: 27) and GenBank Accession number 95P-R76128; human MIP-4, a chemoattractant for leukocytes from Adams, et al. (1997) Patent WO 9634891 (SEQ ID NO: 28) and GenBank Accession number 96P-W07203; pituitary expressed chemokine (PGEC) from Bandman, et al., Patent WO 9616979 (SEQ ID NO: 29) and GenBank Accession number 96P-R95691; and human chemokine HCC-1 from Forsmann, et al. (1998) Patent WO 9741230 (SEQ ID NO: 30) and GenBank Accession number 97P-W38171.

The HCC5 chemokine was discovered through searches and careful analysis of database sequences. The HCC5 sequence was discovered in a cDNA library from pooled bulk breast tumor tissue. Absence of overlapping sequences from other sources suggests extremely specific tissue expression, or highly regulated expression. Amino acid homology analysis suggests that the HCC5 gene encodes a member of a group of related family of chemokines. The primate, e.g., human, HCC5 chemokine is most closely related in sequence to the chemokines, human chemokine HCC1; human pituitary expressed chemokine (PGEC); human MIP-4 (a chemoattractant for leukocytes); human macrophage inflammatory protein-1-gamma (MIP-1γ); and human stem cell mobilizing chemokine (CKbeta-1).

The HCC5 chemokine is seemingly specifically expressed, since its sequence has not appeared from many sources. The structural

similarity to other chemokines suggests that signals important in inflammation, cell differentiation, and development are mediated by it.

It is possible that the HCC5 may actually be an antagonist of one, some, or all, of many related chemokines. In such case, combination compositions may be desired. For example, a combined group of functional agonists and antagonists for specific receptors may be called for, e.g., a combination of chemokines and antibody antagonists of others. In addition, HCC5 may be useful to block HIV or HTLV infection, which viruses may use the respective receptors for infection.

The HCC5 chemokine exhibits limited similarity to portions of known chemokines. See, e.g., Matsushima and Oppenheim (1989) Cytokine 1:2-13; Oppenheim, et al. (1991) Ann. Rev. Immunol. 9:617-648; Schall (1991) Cytokine 3:165-183; and Gronenborn and Clore (1991) Protein Engineering 4:263-269. Other features of comparison are apparent between the HCC5 chemokine and chemokine families. See, e.g., Lodi, et al. (1994) Science 263:1762-1766. In particular, β -sheet and α -helix residues can be determined using, e.g., RASMOL program, see Sayle and Milner-White (1995) TIBS 20:374-376; or Gronenberg, et al. (1991) Protein Engineering 4:263-269; and other structural features are defined in Lodi, et al. (1994) Science 263:1762-1767. These secondary and tertiary features assist in defining further the C, CC, CXC, and CX3C structural features, along with spacing of appropriate cysteine residues.

Antagonists might be created by N-terminal modification, e.g., either truncation or addition of an N-terminal methionine. Since HCC5 is structurally related to the HCC chemokines, it may well exhibit similar behaviors and functions.

The distribution of the HCC5 chemokines, especially in dendritic cells, or in Th1 T cells, B cells, and macrophages, suggest roles in immune functions, e.g., it will likely attract T cells and monocytes. Thus, the HCC5 chemokine is likely to recruit these cell types in vivo, thereby enhancing the immune response mediated by these cell types. The expression patterns

appear consistent with a functional importance of the ligands in a TH1/TH2 regulation and/or response, including, e.g., in a cancer therapy. Thus, ligands and homologs are identified as possible immune adjuvants, e.g., for cellular responses, but also as possible adjuvants to modulate soluble antigen responses, e.g., vaccines.

The invention further provides mammalian, e.g., primate, DNA sequences encoding proteins which exhibit structural properties of likely intracellular deubiquitinating protein enzymes. These proteins are designated deubiquitinating 11 (Dub11) and deubiquitinating 12 (Dub12). For a review of the superfamily of deubiquitinating enzymes see, e.g., Hochstrasser (1995) Curr. Opin. Cell Biol. 7:215-223; Wilkinson, et al. (1995) Biochemistry 34:14535-14546; Baker, et al. (1992) J. Biol. Chem. 267:23364-23375; and Papa and Hochstrasser (1993) Nature 366:313-319. However, the deubiquitinating enzymes have also been reported to have additional functions besides deubiquitination. See, e.g., Hochstrasser (1996) Cell 84:813-815; Hicke and Riezman (1996) Cell 84:277-287; and Chen, et al. (1996) Cell 84:853-862.

The descriptions typically are directed, for exemplary purposes, to the human Dub11 and human Dub12 natural alleles described, but are likewise applicable to allelic and/or polymorphic variants, e.g., from other individuals, as well as splicing variants, e.g., natural forms, and species variants from other primates or other species. These genes will allow isolation of other primate genes encoding proteins related to this, further extending the family beyond the specific embodiments described.

The Dub11 or Dub12 proteins (naturally occurring or recombinant), fragments thereof, and antibodies thereto, along with compounds identified as having binding affinity to Dub11 or Dub12, may be useful in the treatment of conditions associated with abnormal physiology or development, such as, e.g., uterine carcinoma associated with p53 dysregulation associated with human papilloma virus or mental retardation of Angelman syndrome (AS) due to disruption of the 5' end of the UBE3A (E6-AP) gene which codes for a disubiquitination protein. Pharmacological intervention which alters the half-lives of cellular proteins

associated with these diseases may have wide therapeutic potential. Specifically, prevention of p53 ubiquitination (and subsequent degradation) in human papilloma virus positive tumors, and perhaps all tumors retaining wild-type p53 but lacking the retinoblastoma gene function, could lead to programmed cell death. Additionally, specific inhibitors of p27 and cyclin B ubiquitination are predicted to be potent antiproliferative agents. Inhibitors of IkappaB ubiquitination should prevent NFkappaB activation and may have utility in a variety of autoimmune and inflammatory conditions. Finally, deubiquitination enzymes may be novel, potential drug targets as they also appear to regulate cell proliferation. These conditions or disease states may be modulated by appropriate therapeutic treatment using the deubiquitination compositions provided herein.

Conversely, methods for blocking the enzymatic activities should have the opposite effects. Small molecule drug screening to block enzymatic activity of the protein can be performed to identify entities which will block the deubiquitination, thereby affecting protein degradation pathways, as appropriate.

The T cell growth factor interleukin-2 (IL-2) regulates lymphocyte proliferation by inducing the expression of growth promoting genes. HTLV-1 transformed cell lines derived from Adult T-cell Leukemia (ATL) can exhibit constitutive activation of the IL-2-induced JAK/STAT pathway. See Migone, et al. (1998) Proc. Nat'l Acad. Sci. USA 95:3845-3850. ATL cell lines were examined for expression of IL-2 induced genes. It was found that the deubiquitinating enzyme Dub2 is constitutively expressed. See Zhu, et al. (1997) J. Biol. Chem. 272:51-57. Moreover, Dub2 expression conferred cytokine-independent proliferation on the interleukin-3-dependent murine Ba/F3 hematopoietic cell line. SCID mice (n = 18) subcutaneously injected with Ba/F3 cells expressing Dub2, (but not a C to S inactive mutant of Dub2) developed tumors with a six week latency. Cells derived from these tumors exhibited constitutive tyrosine phosphorylation of STAT5 and also mimicked the ATL cell lines by exhibiting down-regulation of the protein tyrosine phosphatase SHP-1. These findings strongly indicate that Dub12 is an oncogene that, when

constitutively expressed, can induce cytokine-independent growth in lymphocytes and may be implicated in leukemogenesis. It is likely that Dub2 controls cell growth by regulating the ubiquitin-dependent proteolysis or the ubiquitin-dependent state of a critical intracellular substrate. Functional similarity of the Dub11 and Dub12 would be expected. Thus, the biological role of Dub2 suggests similar important roles for the other Dub family members.

Screening for inhibitors of the DUB enzymes can also be easily accomplished using the known assays for activity. Such assays can be developed into high throughput screening efforts, testing, particularly, compounds known to affect protein turnover, or similar enzymatic sites. Small molecule antagonists of the enzymes, which will be membrane permeable, would be particularly desirable therapeutically.

In the MD embodiments of the present invention, mammalian, e.g., primate, and rodent, e.g., mouse, DNA sequences are provided encoding proteins which exhibit structural properties of ligands for proteins exhibiting a leucine-rich protein motif (LRR) that is important, e.g., in immune function. These proteins are designated herein human MD-1, human MD-2, and murine MD-2. The human MD-1 is a primate homolog of the previously described rodent MD-1, see, e.g., Miyake, et al. (1998) J. Immunol. 161:1348-1353, while human MD-2 and mouse MD-2 are newly discovered MD-1 homolog. For a general review of LRR proteins, see, e.g., Kobe and Deisenhofer (1994) Trends Biochem. Sci. 19:412. For the role of LRR in specific immune defenses, see, e.g., Jones, et al. (1994) Science 266:789; Dixon, et al. (1996) Cell 84:451; and Baker, et al. (1997) Science 276:726.

Similar sequences for proteins in other species should also be available. The descriptions below are directed, for exemplary purposes, to the primate, e.g., human, MD-1 and MD-2, and rodent, e.g., mouse, MD-2 natural alleles described, but are likewise applicable to allelic and/or polymorphic variants, e.g., from other individuals, as well as splicing variants, e.g., natural forms, and species variants.

The MD-1 or MD-2 proteins (naturally occurring or recombinant), fragments thereof, and antibodies thereto, along with compounds identified as having binding affinity to MD-1 or MD-2, should be useful in the treatment of conditions associated with abnormal physiology or development, such as, e.g., the recognition of specific pathogenic molecules and the activation of B cell physiology. As indicated above, MD-1 and MD-2 exhibit structural motifs characteristic of ligands for the RP105 or BAS-1 surface receptors. Thus, soluble forms, antibodies, or small molecule drugs which disrupt intercellular signaling mediated by these receptors, will find use in modulating cellular response. These responses will be important in normal or abnormal clinical situations.

The matching of the MD and RP105 may also be easily tested. Identification of the counter receptor for the MD-2 may include testing both the RP105 and BAS-1 genes, along with other screening methods, as described. The likely counter receptor structure for the MDs are RP105, BAS-1, and related genes. Associated proteins which bind to these, including the DUB proteins, may be identified using these techniques, among others.

Another aspect of the invention provides members of the cyclin proteins. The cyclins and their partner catalytic subunits, the cyclin-dependent kinases (Cdks), play key roles in the regulation of eukaryotic cell cycle events. See, e.g., Draetta (1994) Curr. Opin. Cell Biol. 6:842-846; Sherr (1994) Cell 79:551-555; and Ohtsubo, et al. (1995) Mol. Cell. Biol. 15:2612-2624. Cyclins were first identified in marine invertebrates on the basis of their dramatic cell cycle periodic expression during meiotic and mitotic divisions.

A large family of cyclins, designated cyclins A-H, bind and activate different Cdks which are serine/threonine kinases essential for cell cycle progression. The timing of the expression of the various cyclins is key in determining at which phase of the cell cycle (S, G₀, G₁, or G₂) their associated Cdk is active. D-type cyclins are synthesized early in G₁ and bind and activate CDK4 and CDK6. Cyclin E-Cdk2 and Cyclin A-Cdk2 complexes form later in G₁ as cells prepare to begin DNA synthesis. Cyclin

B-cdc2 is active during G₂ and mitosis. See, e.g., Lees (1995) Curr. Opin. Cell Biol. 7:773-780.

Other Cyclin-Cdk complex associated proteins are critical for modulation of cyclin activity. Proteins that co-
5 immunoprecipitated with cyclin E were visualized by SDS-PAGE. However, viability of the cyclin E "knockout" mouse, suggested the existence of redundancy. Moreover, work in other species also suggested that a homolog might exist in human.

Cdks can also exert control on cell division and
10 proliferation by phosphorylating specific intracellular target proteins. This phosphorylation event can induce the cellular transition from the G₁ to the S phase of the cell cycle. See, e.g., Strahler, et al. (1992) Biochem. Biophys. Res. Comm. 185:197-203; Brattsand, et al. (1994) Eur. J. Biochem. 220:359-
15 368; and Li, et al. (1996) Cell 85:319-329. Regulation of the cell cycle machinery is important in development and control of cellular proliferation. Misregulation may lead to proliferative disorders, e.g., neoplastic conditions and cancer. See, e.g., Sherr (1998) Science 274:1672-1677.

20 The novel cyclin gene, designated cyclin E2, exhibits about 49% structural identity to the known human cyclin E. See, e.g., Lew, et al. (1991) Cell 66:1197-1206; and NCBI Entrez accession number M74093. The new variant cyclin E2 sequences are provided in SEQ ID NO: 52 and SEQ ID NO: 53. Notable features on the E2
25 sequence include the cyclin box running from about residue 120-254; and a putative phosphorylation site at thr392. The phosphorylation site is trigger in cyclin E for ubiquitin dependent degradation. See Clurman, et al. (1996) Genes and Development 10:1979-1990. Particularly interesting segments
30 include, e.g., from about 93-100; 98-106; 104-113; 107-121; 120-128; 124-134; 131-137; 172-177; 176-185; 189-193; 196-202; 200-210; 218-223; 228-232; 236-242; 240-245; and 248-252.

The structural homology of these genes to identified families suggests related function of these molecules. For example, PGT
35 homologs should function in transport across cell membranes; TNF receptor family antagonists, or agonists, may act as a co-

stimulatory molecule for regulation of T cell mediated cell activation, and may in fact, cause a shift of T helper cell types, e.g., between Th1 and Th2; chemokines have recognized functional properties; intracellular Dubs have been described and the role in oncogenesis established; membrane associated or soluble forms of signaling proteins such as the MDs are well known; and the role of cyclins in cell cycle regulation are recognized. Alternatively, the ligands or binding structures for the cell surface antigens may serve to regulate cell proliferation or development.

For the TNF ligand molecules, they typically modulate cell proliferation, viability, and differentiation. For example, TNF and FAS can kill cells expressing their respective receptors, including fibroblasts, liver cells, and lymphocytes. Some members of this class of ligands exhibit effects on cellular proliferation of cells expressing their respective receptors, e.g., B cells expressing CD40. These effects on proliferation may also effect subsequent differentiation steps, and may lead, directly or indirectly, to changes in cytokine expression profiles.

The members of the TNF ligand family also exhibit costimulation effects, which may also regulate cellular differentiation or apoptosis. Receptor expressing cells may be protected from activation induced cell death (AICD) or apoptosis. For example, CD40 ligand can have effects on T and B lymphocytes.

The embodiments characterized herein are mostly from human, but additional sequences for proteins in other mammalian species, e.g., primates and rodents, will also be available. See below. In particular, with the polypeptide sequences provided, reverse translation, e.g., using the genetic code, software is available, which will indicate what nucleic acid sequences could encode them. See, e.g., MacVector, Oxford Molecular Group Software. Thus, artificial genes, or redundant oligonucleotides may be selected to isolate natural variants or species counterparts.

II. Purified Protein

Primate, e.g., human, DC-PGT polypeptide sequence is shown in SEQ ID NO: 2; primate, e.g., human, HDTEA84 polypeptide sequence is shown in SEQ ID NO: 6; primate, e.g., human, HSLJD37R

polypeptide sequences are shown in SEQ ID NO: 8, 10, and 12; rodent, e.g., murine, RANKL polypeptide sequence is shown in SEQ ID NO: 17; primate forms of RANKL polypeptide sequence are shown in SEQ ID NO: 19, 21, and 23; primate, e.g., human, HCC5 chemokine polypeptide sequence is shown in SEQ ID NO: 25; primate, e.g., human, Dub11 polypeptide sequences are shown in SEQ ID NO: 32 and 34; primate, e.g., human, Dub12 polypeptide sequences are shown in SEQ ID NO: 36 and 38; primate, e.g., human, MD-1 polypeptide sequence is shown in SEQ ID NO: 42; primate, e.g., human, MD-2 polypeptide sequence is shown in SEQ ID NO: 44 and 46; rodent, e.g., mouse, MD-2 polypeptide sequences are shown in SEQ ID NO: 48 and 49; and primate, e.g., human, cyclin E2 is shown in SEQ ID NO: 54.

These amino acid sequences, provided amino to carboxy, are important in providing sequence information in the antigen allowing for distinguishing the protein from other proteins and exemplifying numerous variants. Moreover, the peptide sequences allow preparation of peptides to generate antibodies to recognize such segments, and nucleotide sequences allow preparation of oligonucleotide probes, both of which are strategies for detection or isolation, e.g., cloning, of genes or cDNAs encoding such sequences.

The purified protein, or proteins will typically comprise a plurality of distinct, e.g., nonoverlapping, segments of the specified length. Typically, the plurality will be at least two, more usually at least three, and preferably 5, 7, or even more. While the length minima are provided, longer lengths, of various sizes, may be appropriate, e.g., one of length 7, and two of length 12. Such peptides are useful for generating antibodies by standard methods, as described herein. Synthetic peptides or purified protein can be presented to an immune system to generate a specific binding composition, e.g., monoclonal or polyclonal antibodies. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (Current ed.) Antibodies: A Laboratory Manual Cold Spring Harbor Press.

For example, the specific binding composition could be used for screening of an expression library made from a cell line which

expresses a clone encoding, e.g., a prostaglandin transporter. The screening can be standard staining of surface expressed protein, or by panning. Screening of intracellular expression can also be performed by various staining or immunofluorescence procedures. The binding compositions could be used to affinity purify or sort out cells expressing the protein. The binding compositions may also be useful in determining qualitative and quantitative expression levels of the proteins in various biological samples, including, e.g., cell types or tissues.

- As used herein, the term, e.g., "human DC-PGT", shall encompass, when used in a protein context, a protein having amino acid sequence shown in SEQ ID NO: 2. Significant polypeptide fragments of such a protein should preserve some of the properties, biological or physical, of the full length protein. Other essentially identical or equivalent proteins may be found in other primates or related species. In addition, binding components, e.g., antibodies, typically bind to, e.g., a DC-PGT, with high affinity, e.g., at least about 100 nM, usually better than about 30 nM, preferably better than about 10 nM, and more preferably at better than about 3 nM. Homologous proteins would be found in mammalian species other than human, e.g., primates or rodents. Non-mammalian species should also possess structurally or functionally related genes and proteins, e.g., birds or amphibians. Similar meanings apply in reference to HDTEA84, HSLJD37R, RANKL, HCC5, Dub11, Dub12, MD-1, MD-2, and cyclin E2.

- The term polypeptide, as used herein, includes a significant fragment or segment, and encompasses a stretch of amino acid residues of at least about 8 amino acids, generally at least 10 amino acids, more generally at least 12 amino acids, often at least 14 amino acids, more often at least 16 amino acids, typically at least 18 amino acids, more typically at least 20 amino acids, usually at least 22 amino acids, more usually at least 24 amino acids, preferably at least 26 amino acids, more preferably at least 28 amino acids, and, in particularly preferred embodiments, at least about 30 or more amino acids. The segments may have lengths of at least 37, 45, 53, 61, 70, 80, 90, etc., and often will encompass a plurality of such matching sequences. The

specific ends of such a segment will be at any combinations within the protein. In certain embodiments, there will be a plurality of distinct, e.g., nonoverlapping, segments of the specified length. Typically, the plurality will be at least two, more usually at least three, and preferably 5, 7, or even more. While the length minima are provided, longer lengths, of various sizes, may be appropriate, e.g., one of length 7, and two of length 12.

The term "binding composition" refers to molecules that bind with specificity to the respective protein or polypeptide, e.g., DC-PGT, e.g., in a cell adhesion pairing type fashion, or an antibody-antigen interaction. Other compounds include, e.g., proteins, which specifically associate with DC-PGT, including in a natural physiologically relevant protein-protein interaction, either covalent or non-covalent. The molecule may be a polymer, or chemical reagent. A functional analog may be an antigen with structural modifications, or it may be a molecule which has a molecular shape which interacts with the appropriate binding determinants. The compounds may serve as agonists or antagonists of the binding interaction, see, e.g., Goodman, et al. (eds. 1990) Goodman & Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press.

Substantially pure, in the polypeptide context, typically means that the protein is free from other contaminating proteins, nucleic acids, and other biologicals derived from the original source organism or cell. Purity may be assayed by standard methods, and will ordinarily be at least about 40% pure, more ordinarily at least about 50% pure, generally at least about 60% pure, more generally at least about 70% pure, often at least about 75% pure, more often at least about 80% pure, typically at least about 85% pure, more typically at least about 90% pure, preferably at least about 95% pure, more preferably at least about 98% pure, and in most preferred embodiments, at least 99% pure. The analysis may be weight or molar percentages, evaluated, e.g., by gel staining, spectrophotometry, or terminus labeling. Carriers or excipients will often be subsequently added.

Solubility of a polypeptide or fragment depends upon the environment and the polypeptide. Many parameters affect

polypeptide solubility, including temperature, electrolyte environment, size and molecular characteristics of the polypeptide, and nature of the solvent. Typically, the temperature at which the polypeptide is used ranges from about 4° C to about 65° C. Usually the temperature at use is greater than about 18° C. For diagnostic purposes, the temperature will usually be about room temperature or warmer, but less than the denaturation temperature of components in the assay. For therapeutic purposes, the temperature will usually be body temperature, typically about 37° C for humans and mice, though under certain situations the temperature may be raised or lowered in situ or in vitro.

The size and structure of the polypeptide should generally be in a substantially stable state, and usually not in a denatured state. The polypeptide may be associated with other polypeptides in a quaternary structure, e.g., to confer solubility, or associated with lipids or detergents in a manner which approximates natural lipid bilayer interactions.

The solvent and electrolytes will usually be a biologically compatible buffer, of a type used for preservation of biological activities, and will usually approximate a physiological aqueous solvent. Usually the solvent will have a neutral pH, typically between about 5 and 10, and preferably about 7.5. On some occasions, one or more detergents will be added, typically a mild non-denaturing one, e.g., CHS (cholesteryl hemisuccinate) or CHAPS (3-[3-cholamidopropyl]dimethylammonio]-1-propane sulfonate), or a low enough concentration as to avoid significant disruption of structural or physiological properties of the protein.

Solubility is reflected by sedimentation measured in Svedberg units, which are a measure of the sedimentation velocity of a molecule under particular conditions. The determination of the sedimentation velocity was classically performed in an analytical ultracentrifuge, but is typically now performed in a standard ultracentrifuge. See, Freifelder (1982) Physical Biochemistry (2d ed.), W.H. Freeman; and Cantor and Schimmel (1980) Biophysical Chemistry, parts 1-3, W.H. Freeman & Co., San Francisco; each of which is hereby incorporated herein by reference. As a crude

determination, a sample containing a putatively soluble polypeptide is spun in a standard full sized ultracentrifuge at about 50K rpm for about 10 minutes, and soluble molecules will remain in the supernatant. A soluble particle or polypeptide will typically be less than about 30S, more typically less than about 15S, usually less than about 10S, more usually less than about 6S, and, in particular embodiments, preferably less than about 4S, and more preferably less than about 3S.

The human complimentary DNA and deduced amino acid sequence provided here for DC-PGT contains sequences corresponding to twelve putative transmembrane (TM) segments, based upon a hydropathicity and structural analysis of DC-PGT. A TopPredII (Claros and von Heijne (1994) Comp. Applic. Biosci. 10:685-686) profile of the DC-PGT sequence showing peaks that reach beyond 'putative' or 'certain' baselines. The 12 transmembrane segments correspond to hydrophobic stretches which run approximately from amino acids 47-68 (TM1); 88-107 (TM2); 117-136 (TM3); 188-208 (TM4); 225-244 (TM5); 279-294 (TM6); 367-386 (TM7); 412-431 (TM8); 450-474 (TM9); 561-578 (TM10); 597-616 (TM11); and 651-675 (TM12). Charged amino residues located within the transmembrane domains are: glutamine at amino residues 59, 62, 196, 207, 380, 469, 602, 655, and 675; glutamic acid at residue 95; and arginine at residues 607 and 674. Extracellular loops correspond approximately to amino acid residues 69-87, 137-187, 295-366, 432-449, 579-596, and 617-650. Putative N-glycosylation sites in the exposed, extracellular face of the molecule are located in the second and fifth extracellular loops at Asn-X-Ser/Thr motifs (e.g., 146-148; 176-178; and 538-540). Intracellular portions correspond approximately to amino acid residues 1-46, 108-116, 209-224, 295-366, 432-449, 579-596, and 676-709. These boundaries will often be the boundaries of segments of interest, which be include multiple described segments.

Transporters of this family are typically 12 transmembrane proteins of approximately 650 amino acids in length and include the organic anion transporters in man and rat, prostaglandin transporters of man (Lu, et al. (1996) J. Clin. Invest. 98:1142-1149) and rat; brain digoxin transporters and Matrin F/G of rat

(Kanai, et al. (1995) Science 268:866-869). Characteristic of this family of organic anion transporter proteins is a cysteine rich region located in one of the extracellular loops, which resembles a zinc finger motif. The DC-PGT cysteine rich region is located in extracellular loop 5 with cysteines approximately at positions 489, 493, 495, 504, 516, 520, 539, 541, 554, and 557.

Other particularly interesting segments of the TNF receptors, Dubs, MDs, and cyclin E are pointed out. These may also be segments of comparison with other proteins or genes.

III. Physical Variants

This invention also encompasses proteins or peptides having substantial amino acid sequence homology with the amino acid sequences of the described proteins. The variants include species and polymorphic variants, e.g., naturally occurring forms.

Amino acid sequence homology, or sequence identity, is determined by optimizing residue matches, if necessary, by introducing gaps as required. See also Needleham, et al. (1970) J. Mol. Biol. 48:443-453; Sankoff, et al. (1983) Chapter One in Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA; and the University of Wisconsin Genetics Computer Group, Madison, WI. Sequence identity changes when considering conservative substitutions as matches. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Homologous amino acid sequences are typically intended to include natural polymorphic or allelic and interspecies variations in each respective protein sequence. Typical homologous proteins or peptides will have from 25-100% identity (if gaps can be introduced), to 50-100% identity (if conservative substitutions are included) with the amino acid sequence of the HDTEA84. Identity measures will be at least about 35%, generally at least about 40%, often at least about 50%, typically at least about 60%, usually at least about 70%.

preferably at least about 80%, and more preferably at least about 90%.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When
5 using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s)
10 relative to the reference sequence, based on the designated program parameters.

Optical alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment
15 algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics
20 Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel et al., supra).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and
25 percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) J. Mol. Evol. 35:351-360. The method used is similar to the method described by Higgins and Sharp
30 (1989) CABIOS 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the
35 next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final

alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters.

- 5 For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

- Another example of algorithm that is suitable for determining
10 percent sequence identity and sequence similarity is the BLAST algorithm, which is described Altschul, et al. (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves
15 first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score
20 threshold (Altschul, et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in
25 each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The
30 BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

- 35 In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul

(1993) Proc. Nat'l Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

A further indication that two nucleic acid sequences of polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

The isolated DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub, MD-1, MD-2, or cyclin E2 DNAs can be readily modified by nucleotide substitutions, nucleotide deletions, nucleotide insertions, and inversions of nucleotide stretches. These modifications result in novel DNA sequences which encode these antigens, their derivatives, or proteins having similar physiological, immunogenic, antigenic, or other functional activity. These modified sequences can be used to produce mutant antigens or to enhance expression. Enhanced expression may involve gene amplification, increased transcription, increased translation, and other mechanisms. For example, "Mutant HDTEA84" encompasses a polypeptide otherwise falling within the sequence identity definition of the HDTEA84 as set forth above, but having an amino acid sequence which differs from that of HDTEA84 as normally found in nature, whether by way of deletion, substitution, or insertion. This generally includes proteins having significant identity with a protein having sequence of SEQ ID NO: 6, and as sharing various biological activities, e.g., antigenic or immunogenic, with those

sequences, and in preferred embodiments contain most of the full length disclosed sequences. Full length sequences will typically be preferred, though truncated versions, e.g., soluble constructs and intact domains, will also be useful, likewise, genes or
5 proteins found from natural sources are typically most desired. Similar concepts apply to different HDTEA84 proteins, particularly those found in various warm blooded animals, e.g., mammals and birds, or fish. These descriptions are generally meant to encompass all HDTEA84 proteins, not limited to the particular
10 human embodiment specifically discussed. Similar concepts apply to the other polypeptides provided.

DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub, MD-1, MD-2, or cyclin E2 mutagenesis can also be conducted by making amino acid insertions or deletions. Although site specific mutation sites
15 are predetermined, mutants need not be site specific. Protein mutagenesis can be conducted by making amino acid insertions or deletions, or combinations may be generated to arrive at a final construct. Insertions include amino- or carboxy- terminal fusions. Random mutagenesis can be conducted at a target codon
20 and the expressed mutants can then be screened for the desired activity. Methods for making substitution mutations at predetermined sites in DNA having a known sequence are well known in the art, e.g., by M13 primer mutagenesis or polymerase chain reaction (PCR) techniques. See, e.g., Sambrook, et al. (1989);
25 Ausubel, et al. (1987 and Supplements); and Kunkel, et al. (1987). Methods in Enzymol. 154:367-382.

The mutations in the DNA normally should not place coding sequences out of reading frames and preferably will not create complementary regions that could hybridize to produce secondary
30 mRNA structure such as loops or hairpins.

The present invention also provides recombinant proteins, e.g., heterologous fusion proteins using segments from these proteins. A heterologous fusion protein is a fusion of proteins or segments which are naturally not normally fused in the same
35 manner. Thus, the fusion product of an immunoglobulin with a polypeptide is a continuous protein molecule having sequences fused in a typical peptide linkage, typically made as a single

translation product and exhibiting properties derived from each source peptide. A similar concept applies to heterologous nucleic acid sequences.

5 The present invention also provides recombinant proteins, e.g., heterologous fusion proteins using segments from these proteins. A heterologous fusion protein is a fusion of proteins or segments which are naturally not normally fused in the same manner. A similar concept applies to heterologous nucleic acid sequences. Fusion proteins will be useful as sources for
10 cleaving, separating, and purifying portions thereof.

In addition, new constructs may be made from combining similar functional domains from other proteins. For example, target-binding or other segments may be "swapped" between different new fusion polypeptides or fragments. See, e.g.,
15 Cunningham, et al. (1989) Science 243:1330-1336; and O'Dowd, et al. (1988) J. Biol. Chem. 263:15985-15992.

The phosphoramidite method described by Beaucage and Carruthers (1981) Tetra. Letts. 22:1859-1862, will produce suitable synthetic DNA fragments. A double stranded fragment will
20 often be obtained either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence, e.g., PCR techniques.

25 IV. Functional Variants

The blocking of physiological response with, e.g., HDTEA84, HSLJD37R, RANKL, HCC5 chemokine, MD-1, or MD-2, may result from the inhibition of binding of the respective ligand to signaling form of receptor or binding counterstructure, e.g., through
30 competitive inhibition. In others, binding affinity to substrate may be modifiable or competed with, e.g., DC-PGT, Dubs, or cyclin E2. Thus, in vitro assays of the present invention will often use isolated protein, soluble fragments comprising ligand or substrate binding segments of these proteins, or forms attached to solid
35 phase substrates. These assays will also allow for the diagnostic determination of the effects of either binding segment mutations

and modifications, or antigen mutations and modifications, e.g., HDTEA84, HSLJD37R, RANKL, MD-1, or MD-2 analogs.

This invention also contemplates the use of competitive drug screening assays, e.g., where neutralizing antibodies to antigen or binding fragments compete with a test compound for binding to the protein, e.g., of natural protein sequence. This is applicable to substrate binding, e.g., competitive inhibitors, and in receptor interaction, where the protein has a binding counterstructure.

- 10 "Derivatives" of , e.g., receptor, antigens include amino acid sequence mutants from naturally occurring forms, glycosylation variants, and covalent or aggregate conjugates with other chemical moieties. Covalent derivatives can be prepared by linkage of functionalities to groups which are found in receptor amino acid side chains or at the N- or C- termini, e.g., by standard means. See, e.g., Lundblad and Noyes (1988) Chemical Reagents for Protein Modification, vols. 1-2, CRC Press, Inc., Boca Raton, FL; Hugli (ed. 1989) Techniques in Protein Chemistry, Academic Press, San Diego, CA; and Wong (1991) Chemistry of Protein Conjugation and Cross Linking, CRC Press, Boca Raton, FL.

In particular, glycosylation alterations are included, e.g., made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing, or in further processing steps. See, e.g., Elbein (1987) Ann. Rev. Biochem. 56:497-534.

- 25 Also embraced are versions of the peptides with the same primary amino acid sequence which have other minor modifications, including phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine.

- Fusion polypeptides between these proteins and other homologous or heterologous proteins are also provided. Many cytokine receptors or other surface proteins are multimeric, e.g., homodimeric entities, and a repeat construct may have various advantages, including lessened susceptibility to proteolytic cleavage. Typical examples are fusions of a reporter polypeptide, e.g., luciferase, with a segment or domain of a protein, e.g., a receptor-binding segment, so that the presence or location of the fused ligand may be easily determined. See, e.g., Dull, et al.,

U.S. Patent No. 4,859,609. Other gene fusion partners include bacterial β -galactosidase, trpE, Protein A, β -lactamase, alpha amylase, alcohol dehydrogenase, yeast alpha mating factor, and detection or purification tags such as a FLAG sequence of His6 sequence. See, e.g., Godowski, et al. (1988) Science 241:812-816. Of particular interest are fusion constructs of receptor with a membrane attachment domain.

Fusion peptides will typically be made by either recombinant nucleic acid methods or by synthetic polypeptide methods.

Techniques for nucleic acid manipulation and expression are described generally, e.g., in Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.), vols. 1-3, Cold Spring Harbor Laboratory; and Ausubel, et al. (eds. 1993) Current Protocols in Molecular Biology, Greene and Wiley, NY. Techniques for synthesis of polypeptides are described, e.g., in Merrifield (1963) J. Amer. Chem. Soc. 85:2149-2156; Merrifield (1986) Science 232: 341-347; Atherton, et al. (1989) Solid Phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford; and Grant (1992) Synthetic Peptides: A User's Guide, W.H. Freeman, NY.

This invention also contemplates the use of derivatives of the proteins other than variations in amino acid sequence or glycosylation. Such derivatives may involve covalent or aggregative association with chemical moieties. Covalent or aggregative derivatives will be useful as immunogens, as reagents in immunoassays, or in purification methods such as for affinity purification of binding partners, e.g., other antigens. The desired proteins can be immobilized by covalent bonding to a solid support such as cyanogen bromide-activated SEPHAROSE, by methods which are well known in the art, or adsorbed onto polyolefin surfaces, with or without glutaraldehyde cross-linking, for use in the assay or purification of antibodies or an alternative binding composition. The protein can also be labeled with a detectable group, e.g., for use in diagnostic assays. Purification may be effected by an immobilized antibody or complementary binding partner. Conversely, immunoabsorption or immunodepletion techniques may be developed.

A solubilized protein or fragment of this invention can be used as an immunogen for the production of antisera or antibodies specific for binding to the antigen or fragments thereof. Purified antigen can be used to screen monoclonal antibodies or antigen-binding fragments, encompassing antigen binding fragments of natural antibodies, e.g., Fab, Fab', F(ab)₂, etc. Purified protein can also be used as a reagent to detect antibodies generated in response to the presence of elevated levels of the antigen or cell fragments containing the antigen, both of which may be diagnostic of an abnormal or specific physiological or disease condition. This invention contemplates antibodies raised against amino acid sequences encoded by nucleotide sequences described, or fragments of proteins containing it. In particular, this invention contemplates antibodies having binding affinity to or being raised against specific fragments, e.g., which are predicted to lie outside of the lipid bilayer, both extracellular or intracellular.

The present invention contemplates the isolation of additional closely related species variants. Southern and Northern blot analysis should establish that similar genetic entities exist in other mammals. It is likely that these proteins are widespread in species variants, e.g., rodents, lagomorphs, carnivores, artiodactyla, perissodactyla, and primates.

The invention also provides means to isolate a group of related antigens displaying both distinctness and similarities in structure, expression, and function. Elucidation of many of the physiological effects of the molecules will be greatly accelerated by the isolation and characterization of additional distinct species variants of them. In particular, the present invention provides useful probes for identifying additional homologous genetic entities in different species.

The isolated genes will allow transformation of cells lacking expression of a corresponding protein, e.g., either species types or cells which lack corresponding antigens and exhibit negative background activity. This should allow analysis of the function of genes in comparison to untransformed control cells.

Dissection of critical structural elements which effect the various activation or differentiation functions mediated through these antigens is possible using standard techniques of modern molecular biology, particularly in comparing members of the related class. See, e.g., the homolog-scanning mutagenesis technique described in Cunningham, et al. (1989) Science 243:1339-1336; and approaches used in O'Dowd, et al. (1988) J. Biol. Chem. 263:15985-15992; and Lechleiter, et al. (1990) EMBO J. 9:4381-4390.

The invention also provides, in the context of the DC-PGT, means to isolate a group of related organic anion transporters, e.g., other vertebrate prostaglandin transporters, displaying both distinctness and similarities in structure, expression, and function. Elucidation of many of the physiological effects of the antigens will be greatly accelerated by the isolation and characterization of distinct species variants. In particular, the present invention provides useful probes for identifying additional homologous genetic entities in different species. The results described above indicate that sufficiently homologous genes exist in other species that cross-species hybridization is likely to allow successful cloning.

The isolated genes will allow transformation of cells lacking expression of a described gene, e.g., prostaglandin transporter. Various species types or cells which lack corresponding proteins can be isolated, and should exhibit negative background activity. Expression of transformed genes will allow isolation of antigenically pure cell lines, with defined or single specie variants. This approach will allow for more sensitive detection and discrimination of the physiological effects of the gene, e.g., prostaglandin transporters. Subcellular fragments, e.g., cytoplasts or membrane fragments, can be isolated and used.

The DC-PGT genes may also be useful to increase the rate of transport of desired prostaglandins into transformed cells. Thus, the transporter may be transformed into cells for targeting of incorporation of desired substrates or analogs. For instance, it may be useful to incorporate specific modified prostaglandins into those cells, which may become more susceptible to other

treatments, or directly affected. Thus, specific dendritic cell subsets may be transformed to become more sensitive to prostaglandins or specific substrates. Conversely, such cells may be useful screening targets to identify entities which can block transport, thereby preventing uptake of substrate.

Structural studies of the transporter will lead to design of new variants, particularly analogs exhibiting modified binding affinity, or perhaps, altered rate of transporter activity. This can be combined with previously described screening methods to isolate variants exhibiting desired spectra of activities. Alternatively, many different prostaglandins and analogs thereof may be screened for either transporter binding affinity or transporter transfer. The transporter may require a direct energy source, e.g., ATP or other nucleotide triphosphate, or may depend upon an ion gradient, as described above.

In the context of the Dubs and cyclin E2, intracellular functions would probably involve segments of the antigen which are normally accessible to the cytosol, as would segments of the receptors. However, protein internalization may occur under certain circumstances, and interaction between intracellular components and "extracellular" components may occur. The specific segments of interaction of protein with other intracellular components may be identified by mutagenesis or direct biochemical means, e.g., cross-linking or affinity methods.

Structural analysis by crystallographic or other physical methods will also be applicable. Further investigation of the mechanism of signal transduction will include study of associated components which may be isolatable by affinity methods or by genetic means, e.g., complementation analysis of mutants.

Further study of the expression and control of the proteins will be pursued. The controlling elements associated with the antigens should exhibit differential physiological, developmental, tissue specific, or other expression patterns. Upstream or downstream genetic regions, e.g., control elements, are of interest. In particular, physiological or developmental variants, e.g., multiple alternatively processed forms of the antigen might be found. Thus, differential splicing of message may lead to an

assortment of membrane bound forms, soluble forms, and modified versions of antigen.

Structural studies of the antigens will lead to design of new antigens, particularly analogs exhibiting agonist or antagonist properties on the molecule. This can be combined with previously described screening methods to isolate antigens exhibiting desired spectra of activities.

V. Antibodies

Antibodies can be raised to the various described polypeptides, including species, polymorphic, or allelic variants, and fragments thereof, both in their naturally occurring forms and in their recombinant forms. Additionally, antibodies can be raised to the proteins in either their active forms or in their inactive forms, including native or denatured versions. Anti-idiotypic antibodies are also contemplated.

Antibodies, including binding fragments and single chain versions, against predetermined fragments of the antigens can be raised by immunization of animals with conjugates of the fragments with immunogenic proteins. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptide, or screened for agonistic or antagonistic activity. Antibodies may be agonistic or antagonistic, e.g., by sterically blocking partner or substrate binding. These monoclonal antibodies will usually bind with at least a K_D of about 1 mM, more usually at least about 300 μ M, typically at least about 100 μ M, more typically at least about 30 μ M, preferably at least about 10 μ M, and more preferably at least about 3 μ M or better. More preferred embodiments may have even higher affinities, e.g., at least 300 nM, 30 nM, 3 nM, or perhaps even picomolar affinity.

The term "binding composition" refers to molecules that bind with affinity and selectivity to, e.g., the DC-PGT, e.g., in an antibody-antigen interaction. However, other compounds, e.g., accessory proteins, may also specifically and/or selectively associate with the antigen to the exclusion of other molecules. Typically, the association will be in a natural physiologically

relevant protein-protein interaction, either covalent or non-covalent, and may include members of a multiprotein complex, including carrier compounds or dimerization partners. The molecule may be a polymer, or chemical reagent. No implication as to whether an antigen is necessarily a convex shaped molecule, e.g., the ligand or the receptor of a ligand-receptor interaction, is necessarily represented, other than whether the interaction exhibits similar specificity, e.g., specific or selective affinity. A functional analog may be a polypeptide with structural modifications, e.g., a mutein, or may be a wholly unrelated molecule, e.g., which has a molecular shape which interacts with the appropriate ligand binding determinants. The ligands may serve as agonists or antagonists of the receptor, see, e.g., Goodman, et al. Goodman & Gilman's: The Pharmacological Bases of Therapeutics (current edition) Pergamon Press, Tarrytown, N.Y.

The term "binding agent:antigen complex", as used herein, refers to a complex of a binding agent and antigen, e.g., a DC-PGT protein, that is formed by specific binding of the binding agent to antigen. Specific or selective binding of the binding agent means that the binding agent has a specific binding site, e.g., antigen binding site, that recognizes a site on the antigen. For example, antibodies raised to a DC-PGT protein and recognizing an epitope on the protein are capable of forming a binding agent:DC-PGT protein complex by specific selective binding. Typically, the formation of a binding agent:DC-PGT protein complex allows the qualitative or quantitative measurement of DC-PGT protein in a mixture of other proteins and biologics. The term "antibody:DC-PGT protein complex" refers to an embodiment in which the binding agent, e.g., is the antigen binding portion from an antibody. The antibody may be monoclonal, polyclonal, or a binding fragment of an antibody, e.g., an Fab or F(ab)₂ fragment. The antibody will preferably be a polyclonal antibody for cross-reactivity testing purposes.

The phrase "specifically binds to an antibody" or "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of

the presence of the protein in the presence of a heterogeneous population of other proteins and other biological components. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not significantly bind other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity or selectivity for a particular protein. Often, the serum can be immunoselected or immunodepleted, to minimize crossreactivity with a specific target protein.

A DC-PGT polypeptide that specifically binds to, or that is specifically immunoreactive with, an antibody, e.g., such as a polyclonal antibody, generated against a defined immunogen, e.g., such as an immunogen consisting of an amino acid sequence of SEQ ID NO: 2, or fragments thereof, or a polypeptide generated from the nucleic acid of SEQ ID NO: 1 is typically determined in an immunoassay. Included within the metes and bounds of the present invention are those nucleic acid sequences described herein, including functional variants, that encode polypeptides that selectively bind to polyclonal antibodies generated against the prototypical DC-PGT polypeptide as structurally and functionally defined herein. The immunoassay typically uses a polyclonal antiserum which was raised, e.g., to a protein of SEQ ID NO: 2. This antiserum is selected to have low crossreactivity against appropriate other PGT family members, preferably from the same species, and any such crossreactivity is removed by immunoabsorption prior to use in the immunoassay. Appropriate selective serum preparations can be isolated, and characterized.

The purified protein or defined peptides are useful for generating antibodies by standard methods, as described above. Synthetic peptides or purified protein can be presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual, Cold Spring Harbor Press. Alternatively, the HDTEA84 can be used as a specific binding reagent, and advantage can be taken

of its specificity of binding, much like an antibody would be used.

For example, the specific binding composition could be used for screening of an expression library made from a cell line which
5 expresses an HDTEA84, HSLJD37R, or RANKL. The screening can be standard staining of surface expressed antigen constructs, or by panning. Screening of intracellular expression can also be performed by various staining or immunofluorescence procedures. The binding compositions could be used to affinity purify or sort
10 out cells expressing the protein.

In order to produce antisera for use in an immunoassay, the protein, e.g., of SEQ ID NO: 2, is isolated as described herein. For example, recombinant protein may be produced in a mammalian cell line. An appropriate host, e.g., an inbred strain of mice
15 such as Balb/c, is immunized with the protein of SEQ ID NO: 2 using a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see Harlow and Lane). Alternatively, a substantially full length synthetic peptide derived from the sequences disclosed herein can be used as an
20 immunogen. Polyclonal sera are collected and titrated against the immunogen protein in an immunoassay, e.g., a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross reactivity against other PGT family
25 members, e.g., human or rat PGT, using a competitive binding immunoassay such as the one described in Harlow and Lane, supra, at pages 570-573. Preferably at least two PGT family members are used in this determination in conjunction with the target. These PGT family members can be produced as recombinant proteins and
30 isolated using standard molecular biology and protein chemistry techniques as described herein. Thus, antibody preparations can be identified or produced having desired selectivity or specificity for subsets of PGT family members.

Immunoassays in the competitive binding format can be used
35 for the crossreactivity determinations. For example, the protein of SEQ ID NO: 2 can be immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the

immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to the protein of SEQ ID NO: 2. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorption or immunodepletion with the above-listed proteins.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than twice the amount of the protein of, e.g., SEQ ID NO: 2 that is required, then the second protein is said to specifically bind to an antibody generated to the immunogen.

The antibodies of this invention can also be useful in diagnostic applications. As capture or non-neutralizing antibodies, they can be screened for ability to bind to the antigens without inhibiting binding by a partner. As neutralizing antibodies, they can be useful in competitive binding assays.

They will also be useful in detecting or quantifying a described protein or its binding partners. See, e.g., Chan (ed. 1987) Immunology: A Practical Guide, Academic Press, Orlando, FL; Price and Newman (eds. 1991) Principles and Practice of Immunoassay, Stockton Press, N.Y.; and Ngo (ed. 1988) Nonisotopic Immunoassay, Plenum Press, N.Y. Cross absorptions or depletions and other tests will identify antibodies which exhibit various spectra of specificities, e.g., unique or shared species specificities.

Further, the antibodies, including antigen binding fragments, of this invention can be potent antagonists that bind to the antigen and inhibit functional binding or inhibit the ability of a binding partner to elicit a biological response. They also can be useful as non-neutralizing antibodies and can be coupled to toxins

or radionuclides so that when the antibody binds to antigen, a cell expressing it, e.g., on its surface, is killed. Further, these antibodies can be conjugated to drugs or other therapeutic agents, either directly or indirectly by means of a linker, and may effect drug targeting. They may be labeled for histology evaluation.

Antigen fragments may be joined to other materials, particularly polypeptides, as fused or covalently joined polypeptides to be used as immunogens. An antigen and its fragments may be fused or covalently linked to a variety of immunogens, such as keyhole limpet hemocyanin, bovine serum albumin, tetanus toxoid, etc. See Microbiology, Hoeber Medical Division, Harper and Row, 1969; Landsteiner (1962) Specificity of Serological Reactions, Dover Publications, New York; Williams, et al. (1967) Methods in Immunology and Immunochemistry, vol. 1, Academic Press, New York; and Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press, NY, for descriptions of methods of preparing polyclonal antisera.

In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites, et al. (eds.) Basic and Clinical Immunology (4th ed.), Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.), Academic Press, New York; and particularly in Kohler and Milstein (1975) in Nature 256:495-497, which discusses one method of generating monoclonal antibodies.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides or alternatively to selection of libraries of antibodies in phage or similar vectors. See, Huse, et al. (1989) "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda," Science 246:1275-1281; and Ward, et al. (1989) Nature 341:544-546. The polypeptides and antibodies of the present invention may be used

with or without modification, including chimeric or humanized antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced, see Cabilly, U.S. Patent No. 4,816,567; Moore, et al., U.S. Patent No. 4,642,334; and Queen, et al. (1989) Proc. Nat'l Acad. Sci. USA 86:10029-10033.

The antibodies of this invention can also be used for affinity chromatography in isolating the protein. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a cell lysate may be passed through the column, the column washed, followed by increasing concentrations of a mild denaturant, whereby the purified protein will be released. See, e.g., Wilchek et al. (1984) Meth. Enzymol. 104:3-55.

Antibodies raised against each protein will also be useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to expression of the respective antigens.

VI. Nucleic Acids

The described peptide sequences and the related reagents are useful in detecting, isolating, or identifying a DNA clone encoding, e.g., the DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub11, Dub12, MD-1, MD-2, or cyclin E2 polypeptides, e.g., from a natural source. Typically, the nucleic acids, particularly natural genes, will be useful in isolating a gene from mammal, and similar procedures will be applied to isolate genes from other species, e.g., warm blooded animals, such as birds and mammals. They will

be useful for isolating genes from domestic pets, e.g., dogs and cats, and livestock, e.g., horse, pigs, cattle, sheep, chickens, turkeys, fish, etc. Cross hybridization will allow isolation of respective counterpart genes from other species. A number of
5 different approaches should be available to successfully isolate a suitable nucleic acid clone.

The peptide sequences allow preparation of peptides to generate antibodies to recognize such segments, and various different methods may be used to prepare such peptides. As used
10 herein, e.g., the term prostaglandin transporter shall encompass, when used in a protein context, a protein having an amino acid sequence shown in Table 1, or a significant fragment of such a protein. It also refers to a vertebrate, e.g., mammal, including human, derived polypeptide which exhibits similar biological
15 function, e.g., antigenic, or interacts with prostaglandin transporter specific binding components, e.g., specific antibodies. These binding components, e.g., antibodies, typically bind to a prostaglandin transporter with high affinity, e.g., at least about 100 nM, usually better than about 30 nM, preferably
20 better than about 10 nM, and more preferably at better than about 3 nM. Still higher affinities are possible, e.g., 100 pM, 30 pM, 100 fM, etc.

This invention contemplates use of isolated DNA or fragments of the present invention to encode a structurally related, e.g.,
25 antigenically related, or biologically active protein, e.g., substrate binding or transporting, prostaglandin transporter, TNF receptor-like proteins, chemokine, Dubs, surface receptors, or cell cycle regulatory proteins, or polypeptide fragments thereof. In addition, this invention covers isolated or recombinant DNA
30 which encodes a structurally related or biologically active protein or polypeptide and that is capable of hybridizing under appropriate conditions with the DNA sequences described herein. Said biologically active protein or polypeptide can be an intact antigen, or fragment, and have an amino acid sequence as disclosed
35 in Tables 1-13. Further, this invention covers the use of isolated or recombinant DNA, or fragments thereof, which encode proteins which are homologous to the respective genes or which

were isolated using cDNA encoding the proteins as a probe. Preferably such homologous genes or proteins will be natural forms isolated from other vertebrates, e.g., warm blooded animals, including mammals, such as primates. The isolated DNA can have the respective regulatory sequences in the 5' and 3' flanks, e.g., promoters, enhancers, poly-A addition signals, and others.

An "isolated" nucleic acid is a nucleic acid, e.g., an RNA, DNA, or a mixed polymer, which is substantially separated from other components which naturally accompany a native sequence, e.g., ribosomes, polymerases, and flanking genomic sequences from the originating species. The term embraces a nucleic acid sequence which has been removed from its naturally occurring intracellular environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems. A substantially pure molecule includes once or currently isolated forms of the molecule. Alternatively, a purified species may be separated from host components from a recombinant expression system. Generally, the nucleic acid will be in a vector or fragment less than about 50 kb, usually less than about 30 kb, typically less than about 10 kb, and preferably less than about 6 kb.

An isolated nucleic acid will generally be a homogeneous composition of molecules, but will, in some embodiments, contain minor heterogeneity. This heterogeneity is typically found at the polymer ends or portions not critical to a desired biological function or activity.

The peptide segments can also be used to predict appropriate oligonucleotides to screen a library. The genetic code, e.g., reverse translation, can be used to select appropriate oligonucleotides useful as probes for screening. See, e.g., SEQ ID NO: 1, 5, 7, 9, 11, 16, 18, 20, 22, 24, 31, 33, 35, 37, 41, 43, 47, or 53. In combination with polymerase chain reaction (PCR) techniques, synthetic oligonucleotides will be useful in selecting correct clones from a library. Complementary sequences will also be used as probes, primers, or antisense strands. Various fragments should be particularly useful, e.g., coupled with

anchored vector or poly-A complementary PCR techniques or with complementary DNA of other peptides.

This invention contemplates use of isolated DNA or fragments to encode a biologically active corresponding polypeptide. In addition, this invention covers isolated or recombinant DNA which encodes a biologically active protein or polypeptide which is capable of hybridizing under appropriate conditions with the DNA sequences described herein. Said biologically active protein or polypeptide can be an intact antigen, or fragment, and have an amino acid sequence disclosed in, e.g., SEQ ID NO: 2, 6, 8, 10, 12, 17, 19, 21, 23, 25, 32, 34, 36, 38, 42, 44, 46, 48, 49, or 54. Further, this invention covers the use of isolated or recombinant DNA, or fragments thereof, which encode proteins which are homologous to a described protein or which was isolated using cDNA encoding such protein as a probe. The isolated DNA can have the respective regulatory sequences in the 5' and 3' flanks, e.g., promoters, enhancers, poly-A addition signals, and others.

A "recombinant" nucleic acid is defined either by its method of production or its structure. In reference to its method of production, e.g., a product made by a process, the process is use of recombinant nucleic acid techniques, e.g., involving human intervention in the nucleotide sequence, typically selection or production. Alternatively, it can be a nucleic acid made by generating a sequence comprising fusion of two fragments which are not naturally contiguous to each other, but is meant to exclude products of nature, e.g., naturally occurring mutants. Thus, e.g., products made by transforming cells with any unnaturally occurring vector is encompassed, as are nucleic acids comprising sequence derived using any synthetic oligonucleotide process. Such is often done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site.

Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a single genetic entity comprising a desired combination of functions not found in the commonly available natural forms. Restriction enzyme recognition sites are often the target of such artificial manipulations, but

other site specific targets, e.g., promoters, DNA replication sites, regulation sequences, control sequences, or other useful features may be incorporated by design. A similar concept is intended for a recombinant, e.g., fusion, polypeptide.

- 5 Specifically included are synthetic nucleic acids which, by genetic code redundancy, encode polypeptides similar to fragments of these antigens, and fusions of sequences from various different species variants.

10 A significant "fragment" in a nucleic acid context is a contiguous segment of at least about 17 nucleotides, generally at least about 22 nucleotides, ordinarily at least about 29 nucleotides, more often at least about 35 nucleotides, typically at least about 41 nucleotides, usually at least about 47 nucleotides, preferably at least about 55 nucleotides, and in
15 particularly preferred embodiments will be at least about 60 or more nucleotides.

A DNA which codes for a DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub, MD-1, MD-2, or cyclin E2 protein will be particularly useful to identify genes, mRNA, and cDNA species which code for
20 related or homologous proteins, as well as DNAs which code for homologous proteins from different species. There are likely homologs in other species, including primates, rodents, birds, and fish. Various such proteins should be homologous and are encompassed herein. However, even genes encoding proteins that
25 have a more distant evolutionary relationship to the antigen can readily be isolated under appropriate conditions using these sequences if they are sufficiently homologous. Primate proteins are of particular interest.

Recombinant clones derived from the genomic sequences, e.g.,
30 containing introns, will be useful for transgenic studies, including, e.g., transgenic cells and organisms, and for gene therapy. See, e.g., Goodnow (1992) "Transgenic Animals" in Roitt (ed.) Encyclopedia of Immunology, Academic Press, San Diego, pp. 1502-1504; Travis (1992) Science 256:1392-1394; Kuhn, et al.
35 (1991) Science 254:707-710; Capecchi (1989) Science 244:1288; Robertson (1987 ed.) Teratocarcinomas and Embryonic Stem Cells: A

Practical Approach, IRL Press, Oxford; and Rosenberg (1992) J. Clinical Oncology 10:180-199.

Substantial homology in the nucleic acid sequence comparison context means either that the segments, or their complementary strands, when compared, are identical when optimally aligned, with appropriate nucleotide insertions or deletions, in at least about 50% of the nucleotides, generally at least about 58%, ordinarily at least about 65%, often at least about 71%, typically at least about 77%, usually at least about 85%, preferably at least about 95 to 98% or more, and in particular embodiments, as high as about 99% or more of the nucleotides. Alternatively, substantial homology exists when the segments will hybridize under selective hybridization conditions, to a strand, or its complement, typically using a sequence of DC-PGT, e.g., in SEQ ID NO: 1.

Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 30 nucleotides, preferably at least about 75% over a stretch of about 25 nucleotides, and most preferably at least about 90% over about 20 nucleotides. See, Kanehisa (1984) Nuc. Acids Res. 12:203-213.

The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will be over a stretch of at least about 17 nucleotides, usually at least about 28 nucleotides, typically at least about 40 nucleotides, and preferably at least about 75 to 100 or more nucleotides.

Stringent conditions, in referring to homology in the hybridization context, will be stringent combined conditions of salt, temperature, organic solvents, and other parameters, typically those controlled in hybridization reactions. Stringent temperature conditions will usually include temperatures in excess of about 30° C, usually in excess of about 37° C, typically in excess of about 55° C, preferably in excess of about 70° C. Stringent salt conditions will ordinarily be less than about 1000 mM, usually less than about 400 mM, typically less than about 250 mM, preferably less than about 150 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson (1968) J. Mol. Biol. 31:349-370. Hybridization under stringent conditions should

give a background of at least 2-fold over background, preferably at least 3-5 or more.

DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub11, Dub12, MD-1, MD-2, or cyclin E2 from other mammalian species can be cloned and isolated by cross-species hybridization of closely related species. Homology may be relatively low between distantly related species, and thus hybridization of relatively closely related species is advisable. Alternatively, preparation of an antibody preparation which exhibits less species specificity may be useful in expression cloning approaches.

VII. Making Proteins; Mimetics

Nucleic acids which encodes the described proteins, or fragments thereof, can be obtained by chemical synthesis, screening cDNA libraries, or screening genomic libraries prepared from a wide variety of cell lines or tissue samples. See, e.g., Okayama and Berg (1982) Mol. Cell. Biol. 2:161-170; Gubler and Hoffman (1983) Gene 25:263-269; and Glover (ed. 1984) DNA Cloning: A Practical Approach, IRL Press, Oxford. Alternatively, the sequences provided herein provide useful PCR primers or allow synthetic or other preparation of suitable genes encoding a receptor; including, naturally occurring embodiments.

DNA can be expressed in a wide variety of host cells for the synthesis of a full-length protein, or fragments, which can in turn, e.g., be used to generate polyclonal or monoclonal antibodies; for binding studies; for construction and expression of modified molecules; for structure/function studies; and for controls in detection assays. Each antigen or its fragments can be expressed in host cells that are transformed or transfected with appropriate expression vectors. These molecules can be substantially purified to be free of protein or cellular contaminants, other than those derived from the recombinant host, and therefore are particularly useful in pharmaceutical compositions when combined with a pharmaceutically acceptable carrier and/or diluent. The antigen, or portions thereof, may be expressed as fusions with other proteins.

Vectors, as used herein, comprise plasmids, viruses, bacteriophage, integratable DNA fragments, and other vehicles which enable the integration of DNA fragments into the genome of the host. See, e.g., Pouwels, et al. (1985 and Supplements)

5 Cloning Vectors: A Laboratory Manual, Elsevier, N.Y.; and Rodriguez, et al. (1988 eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworth, Boston, MA.

Expression vectors are typically self-replicating DNA or RNA constructs containing the desired antigen gene or its fragments, usually operably linked to suitable genetic control elements that are recognized in a suitable host cell. These control elements are capable of effecting expression within a suitable host. The specific type of control elements necessary to effect expression will depend upon the eventual host cell used. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system, and typically include a transcriptional promoter, an optional operator to control the onset of transcription, transcription enhancers to elevate the level of mRNA expression, a sequence that encodes a suitable ribosome binding site, and sequences that terminate transcription and translation. Expression vectors also usually contain an origin of replication that allows the vector to replicate independently of the host cell.

For purposes of this invention, DNA sequences are operably linked when they are functionally related to each other. For example, DNA for a presequence or secretory leader is operably linked to a polypeptide if it is expressed as a preprotein or participates in directing the polypeptide to the cell membrane or in secretion of the polypeptide. A promoter is operably linked to a coding sequence if it controls the transcription of the polypeptide; a ribosome binding site is operably linked to a coding sequence if it is positioned to permit translation. Usually, operably linked means contiguous and in reading frame, however, certain genetic elements such as repressor genes are not contiguously linked but still bind to operator sequences that in turn control expression. See e.g., Rodriguez, et al., Chapter 10, pp. 205-236; Balbas and Bolivar (1990) Methods in Enzymol. 185:14-

37; and Ausubel, et al. (1993) Current Protocols in Molecular Biology, Greene and Wiley, NY.

Representative examples of suitable expression vectors include pCDNA1; pCD, see Okayama, et al. (1985) Mol. Cell Biol. 5:1136-1142; pMC1neo Poly-A, see Thomas, et al. (1987) Cell 51:503-512; and a baculovirus vector such as pAC 373 or pAC 610. See, e.g., Miller (1988) Ann. Rev. Microbiol. 42:177-199.

Usually, expression vectors are designed for stable replication in their host cells or for amplification to greatly increase the total number of copies of the desirable gene per cell. It is not always necessary to require that an expression vector replicate in a host cell, e.g., it is possible to effect transient expression of the antigen or its fragments in various hosts using vectors that do not contain a replication origin that is recognized by the host cell. It is also possible to use vectors that cause integration of a gene or its fragments into the host DNA by recombination, or to integrate a promoter which controls expression of an endogenous gene.

Adenovirus techniques are available for expression of the genes in various cells and organs. See, e.g., Hitt, et al. (1997) Adv. Pharmacol. 40:137-195; and literature from Quantum Biotechnologies, Montreal, Canada. Animals may be useful to determine the effects of the gene on various developmental or physiologically functional animal systems.

Suitable host cells include prokaryotes, lower eukaryotes, and higher eukaryotes. Prokaryotes include both gram negative and gram positive organisms, e.g., *E. coli* and *B. subtilis*. Lower eukaryotes include yeasts, e.g., *S. cerevisiae* and *Pichia*, and species of the genus *Dictyostelium*. Higher eukaryotes include established tissue culture cell lines from animal cells, both of non-mammalian origin, e.g., insect cells, and birds, and of mammalian origin, e.g., human, primates, and rodents.

Prokaryotic host-vector systems include a wide variety of vectors for many different species. As used herein, *E. coli* and its vectors will be used generically to include equivalent vectors used in other prokaryotes. A representative vector for amplifying DNA is pBR322 or many of its derivatives. Vectors that can be

used to express the prostaglandin transporter or its fragments include, but are not limited to, such vectors as those containing the lac promoter (pUC-series); trp promoter (pBR322-trp); Ipp promoter (the pIN-series); lambda-pP or pR promoters (pOTS); or
5 hybrid promoters such as ptac (pDR540). See Brosius et al. (1988) "Expression Vectors Employing Lambda-, trp-, lac-, and Ipp-derived Promoters", in Rodriguez and Denhardt (eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Buttersworth, Boston, Chapter 10, pp. 205-236, which is incorporated herein by
10 reference.

Lower eukaryotes, e.g., yeasts and Dictyostelium, may be transformed with vectors encoding vertebrate prostaglandin transporters. For purposes of this invention, the most common lower eukaryotic host is the baker's yeast, *Saccharomyces*
15 *cerevisiae*. It will be used to generically represent lower eukaryotes although a number of other strains and species are also available. Yeast vectors typically consist of a replication origin (unless of the integrating type), a selection gene, a promoter, DNA encoding the desired protein or its fragments, and
20 sequences for translation termination, polyadenylation, and transcription termination. Suitable expression vectors for yeast include such constitutive promoters as 3-phosphoglycerate kinase and various other glycolytic enzyme gene promoters or such inducible promoters as the alcohol dehydrogenase 2 promoter or
25 metallothionine promoter. Suitable vectors include derivatives of the following types: self-replicating low copy number (such as the YRp-series), self-replicating high copy number (such as the YEpl-series); integrating types (such as the YIp-series), or mini-chromosomes (such as the YCp-series).

30 Higher eukaryotic tissue culture cells are the preferred host cells for expression of the functionally active prostaglandin transporter. In principle, most higher eukaryotic tissue culture cell lines are workable, e.g., insect baculovirus expression systems, whether from an invertebrate or vertebrate source.
35 However, mammalian cells are preferred, in that the processing, both cotranslationally and posttranslationally is more likely to simulate natural forms. Transformation or transfection and

propagation of such cells has become a routine procedure. Examples of useful cell lines include HeLa cells, Chinese hamster ovary (CHO) cell lines, baby rat kidney (BRK) cell lines, insect cell lines, bird cell lines, and monkey (COS) cell lines.

- 5 Expression vectors for such cell lines usually include an origin of replication, a promoter, a translation initiation site, RNA splice sites (if genomic DNA is used), a polyadenylation site, and a transcription termination site. These vectors also usually contain a selection gene or amplification gene. Suitable
10 expression vectors may be plasmids, viruses, or retroviruses carrying promoters derived, e.g., from such sources as from adenovirus, SV40, parvoviruses, vaccinia virus, or cytomegalovirus. Representative examples of suitable expression vectors include pCDNA1; pCD, see Okayama et al. (1985) Mol. Cell
15 Biol. 5:1136-1142; pMC1neo Poly-A, see Thomas et al. (1987) Cell 51:503-512; and a baculovirus vector such as pAC 373 or pAC 610.

- It will often be desired to express a DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub11, Dub12, MD-1, MD-2, or cyclin E2 polypeptide in a system which provides a specific or defined
20 glycosylation pattern. See, e.g., Luckow and Summers (1988) Bio/Technology 6:47-55; and Kaufman (1990) Meth. Enzymol. 185:487-511. Preferred prokaryotic forms lack eukaryotic glycosylation patterns. However, the pattern will be modifiable by exposing the polypeptide, e.g., an unglycosylated form, to appropriate
25 glycosylating proteins introduced into a heterologous expression system. For example, the desired gene may be cotransformed with one or more genes encoding mammalian or other glycosylating enzymes. Using this approach, certain mammalian glycosylation patterns will be achievable or approximated in prokaryote or other
30 cells.

- The DC-PGT, HDTEA84, HSLJD37R, RANKL, HCC5, Dub11, Dub12, MD-1, MD-2, or cyclin E2, or a fragment thereof, may be engineered to be phosphatidyl inositol (PI) linked to a cell membrane, but can be removed from membranes by treatment with a phosphatidyl
35 inositol cleaving enzyme, e.g., phosphatidyl inositol phospholipase-C. This releases the antigen in a biologically active form, and allows purification by standard procedures of

protein chemistry. See, e.g., Low (1989) Biochim. Biophys. Acta 988:427-454; Tse, et al. (1985) Science 230:1003-1008; and Brunner, et al. (1991) J. Cell Biol. 114:1275-1283.

Transformed cells include cells, preferably mammalian, that
5 have been transformed or transfected with vectors containing a
prostaglandin transporter gene, typically constructed using
recombinant DNA techniques. Transformed host cells usually
express the antigen or its fragments, but for purposes of cloning,
amplifying, and manipulating its DNA, do not need to express the
10 protein. This invention further contemplates culturing
transformed cells in a nutrient medium, thus permitting the
protein, or soluble fragments, to accumulate in the culture.
Soluble protein can be recovered, either from the culture or from
the culture medium, and membrane associated proteins may be
15 prepared from suitable cell subfractions.

Now that the genes have been characterized, fragments or
derivatives thereof can be prepared by conventional processes for
synthesizing peptides. These include processes such as are
described in Stewart and Young (1984) Solid Phase Peptide
20 Synthesis, Pierce Chemical Co., Rockford, IL; Bodanszky and
Bodanszky (1984) The Practice of Peptide Synthesis, Springer-
Verlag, New York; and Bodanszky (1984) The Principles of Peptide
Synthesis, Springer-Verlag, New York. For example, an azide
process, an acid chloride process, an acid anhydride process, a
25 mixed anhydride process, an active ester process (for example, p-
nitrophenyl ester, N-hydroxysuccinimide ester, or cyanomethyl
ester), a carbodiimidazole process, an oxidative-reductive
process, or a dicyclohexylcarbodiimide (DCCD)/additive process can
be used. Solid phase and solution phase syntheses are both
30 applicable to the foregoing processes.

The proteins, fragments, or derivatives are suitably prepared
in accordance with the above processes as typically employed in
peptide synthesis, generally either by a so-called stepwise
process which comprises condensing an amino acid to the terminal
35 amino acid, one by one in sequence, or by coupling peptide
fragments to the terminal amino acid. Amino groups that are not

being used in the coupling reaction are typically protected to prevent coupling at an incorrect location.

If a solid phase synthesis is adopted, the C-terminal amino acid is bound to an insoluble carrier or support through its carboxyl group. The insoluble carrier is not particularly limited as long as it has a binding capability to a reactive carboxyl group. Examples of such insoluble carriers include halomethyl resins, such as chloromethyl resin or bromomethyl resin, hydroxymethyl resins, phenol resins, tert-alkyloxycarbonyl-hydrazidated resins, and the like.

An amino group-protected amino acid is bound in sequence through condensation of its activated carboxyl group and the reactive amino group of the previously formed peptide or chain, to synthesize the peptide step by step. After synthesizing the complete sequence, the peptide is split off from the insoluble carrier to produce the peptide. This solid-phase approach is generally described by Merrifield et al. (1963) in J. Am. Chem. Soc. 85:2149-2156, which is incorporated herein by reference.

The prepared protein and fragments thereof can be isolated and purified from the reaction mixture by means of peptide separation, e.g., by extraction, precipitation, electrophoresis and various forms of chromatography, and the like. The proteins of this invention can be obtained in varying degrees of purity depending upon its desired use. Purification can be accomplished by use of the protein purification techniques disclosed herein or by the use of the antibodies herein described in immunoabsorbent affinity chromatography. This immunoabsorbent affinity chromatography is carried out by first linking the antibodies to a solid support and then contacting the linked antibodies with solubilized lysates of appropriate source cells, lysates of other cells expressing the protein, or lysates or supernatants of cells producing the desired protein as a result of DNA techniques, see below. Detergents may be necessary to include in the methods to maintain protein solubility.

VIII. Uses

The present invention provides reagents which will find use in diagnostic applications as described elsewhere herein, e.g., in the general description for cell mediated conditions, or below in the description of kits for diagnosis. The genes will be useful in forensic analyses, e.g., to identify species, or to diagnose different cell subsets or types.

If DC-PGT is used to clear prostaglandins (PGs) and other metabolically active organic anions from the body (in the liver, fetal liver, lung and placenta) it is easy to suppose that an alteration in the capacity of this mechanism could augment the allergic response. Prostaglandin $\text{PGF}_2\alpha$ and PGD_2 , and PGG_2 and thromboxane A_2 can cause airway obstruction, particularly in the peripheral lung, while PGE_2 and PGI_2 are bronchodilators. Use of the transporter of the invention could help transport or remove these prostaglandins to modulate airway obstruction.

Additionally, prostaglandins play an important role in secondary immunosuppression seen following surgical stress. Alexander (1990) J. Trauma 30:S70; Faist, et al. (1987) J. Trauma 27:837; Ninneman, et al. (1984) J. Trauma 24:201; Wood, et al. (1987) Arch. Surg. 122:179; Polk, et al. in Eremin and Sewell (eds. 1992) The Immunological Basis of Surgical Science and Practice, Oxford U. Press. In particular, PGE_2 inhibits lymphocyte proliferation, decreases IL-2 release, decreases response to IL-2, inhibits natural killer cells, and activates suppressor cells. Major injury has been shown to result in increased production of PGE_2 from inhibitory macrophages, with a resulting decrease in production of IL-1 and IL-2. This effect may persist for 7 to 10 days after major injury. Other studies have shown no increase in circulating PGE_2 following burns but did find increased local production with increased sensitivity of lymphocytes to the action of PGE_2 .

Prostaglandin E_2 , through locally produced vasodilatory effects, may play a role in rheumatoid arthritis by promoting the entry of inflammatory cells into the joint. Once in the synovial fluid, polymorphonuclear leukocytes can ingest immune complexes, which, in turn, cause neutrophils to produce reactive oxygen

metabolites and other inflammatory mediators to further enhance an inflammatory cascade. Henson, et al. (1987) J. Clin. Invest. 79:699.

Accordingly, it is possible to use the present invention to
5 modulate prostaglandins in a subject suffering from trauma, injury, disease or in post-surgical treatments.

Immune system cells may be synthesizing PGs and thus using DC-PGT in an efflux role for removing PGs from the intracellular space may be useful. Equally, DC-PGT might transport a specific
10 organic anion. Abnormal proliferation, regeneration, degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal function of a prostaglandin transporter should be a likely target for a
15 substrate or blocking substrate. Alternatively, the transporter may be a useful means for supplying important metabolites or metabolite blockers to the respective cells.

For example, transformation with the transporter may increase availability of the substrate to the cell. In certain situations,
20 a prostaglandin analog might be advantageously supplied to the cell. The prostaglandin analog might confer high susceptibility to further treatment, e.g., radiation sensitivity or otherwise, or may directly affect normal metabolism, e.g., nucleic acid related enzymes. Alternatively, the transporter may be useful to screen
25 for antagonists or inhibitors, which might be effective in blocking the normal availability to the cell of the natural substrate. Screening methods for such prostaglandin analogs are provided.

Screening using prostaglandin transporter for binding
30 metabolites or compounds having binding affinity to the transporter can be performed, including isolation of associated components. Subsequent biological assays can then be utilized to determine if the compound has intrinsic biological activity and is therefore an agonist or antagonist in that it blocks an activity
35 of the transporter. In particular, prostaglandin analogs may be useful in blocking binding of the natural target or otherwise blocking transporter activity. Alternatively, various other

analogs may be useful in blocking an ion transporter, or organic anion source. This invention further contemplates the therapeutic use of antibodies to prostaglandin transporter as antagonists. This approach should be particularly useful with other
5 prostaglandin transporter species variants and other members of the family.

Antagonists of the transporter activity, e.g., antibodies which block the transport, may be useful in various medical conditions. These would include immune, inflammatory or allergic
10 abnormalities, all of which are important where transfer of organic anions take place. Certain congenital diseases of prostaglandin physiology will be susceptible to such a therapeutic approach.

The HDTEA84, HSLJD37R, RANKL, HCC5, MD-1, or MD-2 (naturally
15 occurring or recombinant), fragments thereof, and antibodies thereto, along with compounds identified as having binding affinity to them, should be useful in the treatment of conditions associated with abnormal physiology or development, including abnormal proliferation, e.g., cancerous conditions, or
20 degenerative conditions. In particular, modulation of development of lymphoid cells will be achieved by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal expression or abnormal signaling by a ligand or receptor should be a likely
25 target for an agonist or antagonist of the antigen. The antigen plays a role in regulation or development of hematopoietic cells, e.g., lymphoid cells, which affect immunological responses, e.g., autoimmune disorders.

In particular, the antigen may provide a costimulatory signal
30 to cell activation, or be involved in regulation of cell proliferation or differentiation. Thus, the HDTEA84, HSLJD37R, RANKL, HCC5, MD-1, or MD-2 will likely modulate cells which possess a receptor therefor, e.g., T cell mediated interactions with other cell types. These interactions would lead, in
35 particular contexts, to modulation of cell growth, cytokine synthesis by those or other cells, or development of particular effector cells.

Moreover, the HDTEA84, HSLJD37R, RANKL, HCC5, MD-1, or MD-2 or antagonists could redirect T cell responses, e.g., between Th1 and Th2 polarization, or with Th0 cells, or may affect B cells or other lymphoid cell subsets. Among these agonists should be various antibodies which recognize the appropriate epitopes, e.g., which mimic binding of ligand or receptor to its partner. Alternatively, they may bind to epitopes which sterically can block receptor binding. Bone morphogenesis may be regulated by these receptor segments.

The ligands or receptors may provide a selective and powerful way to modulate immune responses in abnormal situations, e.g., autoimmune disorders, including rheumatoid arthritis, systemic lupus erythematosus (SLE), Hashimoto's autoimmune thyroiditis, as well as acute and chronic inflammatory responses in which T cell activation, expansion, and/or immunological T cell memory play an important role. See also Samter, et al. (eds) Immunological Diseases vols. 1 and 2, Little, Brown and Co. Regulation of bone morphogenesis, T cell activation, expansion, and/or cytokine release by the naturally occurring secreted form of HDTEA84, HSLJD37R, RANKL, HCC5, MD-1, or MD-2, or an antagonist thereof, may be effected.

In addition, certain combination compositions with other modulators of signaling would be useful, especially with the TNF receptor-like genes. Such other signaling molecules might include, e.g., TCR reagents, CD40, CD40L, CTLA-8, CD28, SLAM, FAS, osteoprotegerin, and their respective antagonists, including antibodies.

Cyclin E2 nucleotides, e.g., human cyclin E2 DNA or RNA, may be used as a component in a forensic assay. For instance, the nucleotide sequences provided may be labeled using, e.g., ³²P or biotin and used to probe standard restriction fragment polymorphism blots, providing a measurable character to aid in distinguishing between individuals. Such probes may be used in well-known forensic techniques such as genetic fingerprinting. In addition, nucleotide probes made from cyclin E2 sequences may be used in in situ assays to detect chromosomal abnormalities. For instance, rearrangements in the human chromosome encoding a cyclin

E2 gene may be detected via well-known in situ techniques, using cyclin E2 probes in conjunction with other known chromosome markers. The cyclin E2 gene may have useful prognostic utility in various cancers, e.g., breast, etc.

5 Antibodies and other binding agents directed towards cyclin E2 proteins or nucleic acids may be used to purify the corresponding cyclin E2 molecule. As described in the Examples below, antibody purification of cyclin E2 protein components is both possible and practicable. Antibodies and other binding
10 agents may also be used in a diagnostic fashion to determine whether cyclin E2 protein components are present in a tissue sample or cell population using well-known techniques described herein. The ability to attach a binding agent to a cyclin E2 protein provides a means to diagnose disorders associated with
15 cyclin E2 protein misregulation. Antibodies and other cyclin E2 protein binding agents may also be useful as histological markers. As described in the examples below, cyclin E2 protein expression is limited to specific tissue types. By directing a probe, such as an antibody or nucleic acid to a cyclin E2 protein it is
20 possible to use the probe to distinguish tissue and cell types in situ or in vitro.

 This invention also provides reagents with significant therapeutic value. The cyclin E2 protein (naturally occurring or recombinant), fragments thereof, and antibodies thereto, along
25 with compounds identified as having binding affinity to a cyclin E2 protein, can be useful in the treatment of conditions associated with abnormal physiology or development, including abnormal proliferation, e.g., cancerous conditions, or degenerative conditions. Abnormal proliferation, regeneration,
30 degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal expression or abnormal signaling by a cyclin E2 protein is a target for an agonist or antagonist of the protein. The proteins likely play a
35 role in regulation or development of neuronal or hematopoietic cells, e.g., lymphoid cells, which affect immunological responses.

Various abnormal conditions are known in each of the cell types shown to possess, e.g., HDTEA84, mRNA by Northern blot analysis. See Berkow (ed.) The Merck Manual of Diagnosis and Therapy, Merck & Co., Rahway, NJ; Thorn, et al. Harrison's Principles of Internal Medicine, McGraw-Hill, NY; and Weatherall, et al. (eds.) Oxford Textbook of Medicine, Oxford University Press, Oxford. Many other medical conditions and diseases involve T cells or are T cell mediated, and many of these may be responsive to treatment by an agonist or antagonist provided herein. See, e.g., Stites and Terr (eds; 1991) Basic and Clinical Immunology Appleton and Lange, Norwalk, CT; and Samter, et al. (eds) Immunological Diseases Little, Brown and Co. These problems should be susceptible to prevention or treatment using compositions provided herein.

Specific, or selective, antibodies can be purified and then administered to a patient, veterinary or human. These reagents can be combined for therapeutic use with additional active or inert ingredients, e.g., in conventional pharmaceutically acceptable carriers or diluents, e.g., immunogenic adjuvants, along with physiologically innocuous stabilizers, excipients, or preservatives. These combinations can be sterile filtered and placed into dosage forms as by lyophilization in dosage vials or storage in stabilized aqueous preparations. This invention also contemplates use of antibodies or binding fragments thereof, including forms which are not complement binding.

Drug screening using proteins or fragments thereof can be performed to identify compounds having binding affinity to or other relevant biological effects on antigen functions, including isolation of associated components. Subsequent biological assays can then be utilized to determine if the compound has intrinsic stimulating activity or is a blocker or antagonist in that it blocks the activity of the antigen, e.g., mutein antagonists. Likewise, a compound having intrinsic stimulating activity can activate the signal pathway and is thus an agonist in that it overcomes any blocking activity of these soluble forms of receptors. This invention further contemplates the therapeutic use of blocking antibodies to ligands or receptors as agonists or

antagonists and of stimulatory molecules, e.g., muteins, as agonists. This approach should be particularly useful with other soluble receptor species variants.

Another therapeutic approach included within the invention
5 involves direct administration of reagents or compositions by any conventional administration techniques (e.g., but not restricted to local injection, inhalation, or administered systemically), to the subject with an immune, allergic, or trauma disorder. The reagents, formulations, or compositions included within the bounds
10 and metes of the invention may also be targeted to specific cells or transporters by methods described herein. The actual dosage of reagent, formulation, or composition that modulates an immune, allergic, or trauma disorder depends on many factors, including the size and health of an organism, however one of ordinary skill
15 in the art can use the following teachings describing the methods and techniques for determining clinical dosages. See, e.g., Spilker (1984) Guide to Clinical Studies and Developing Protocols, Raven Press Books, Ltd., New York, pp. 7-13, 54-60; Spilker (1991) Guide to Clinical Trials, Raven Press, Ltd., New York, pp. 93-101;
20 Craig and Stitzel (eds. 1986) Modern Pharmacology, 2d ed., Little, Brown and Co., Boston, pp. 127-33; Speight (ed. 1987) Avery's Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics, 3d ed., Williams and Wilkins, Baltimore, pp. 50-56; Tallarida, et al. (1988) Principles in General Pharmacology,
25 Springer-Verlag, New York, pp. 18-20). Generally, the dose will be in the range of about between 0.5 fg/ml and 500 µg/ml, inclusive, final concentration administered per day to an adult in a pharmaceutically acceptable carrier.

The quantities of reagents necessary for effective therapy
30 will depend upon many different factors, including means of administration, target site, physiological state of the patient, and other medicants administered. Thus, treatment dosages should be titrated to optimize safety and efficacy. Typically, dosages used in vitro may provide useful guidance in the amounts useful
35 for in situ administration of these reagents. Animal testing of effective doses for treatment of particular disorders will provide

further predictive indication of human dosage. Various considerations are described, e.g., in Gilman, et al. (eds. 1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences, 17th ed. (1990), Mack Publishing Co., Easton, Penn. Methods for administration are discussed, e.g., for oral, intravenous, intraperitoneal, or intramuscular administration, transdermal diffusion, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, and other compounds described, e.g., in the Merck Index, Merck & Co., Rahway, New Jersey. Dosage ranges would ordinarily be expected to be in amounts lower than 1 mM concentrations, typically less than about 10 μ M concentrations, usually less than about 100 nM, preferably less than about 10 pM (picomolar), and most preferably less than about 1 fM (femtomolar), with an appropriate carrier. Slow release formulations, or a slow release apparatus will often be utilized for continuous or long term administration. See, e.g., Langer (1990) Science 249:1527-1533.

Ligands, receptors, enzymes, fragments thereof, and antibodies to it or its fragments, antagonists, and agonists, may be administered directly to the host to be treated or, depending on the size of the compounds, it may be desirable to conjugate them to carrier proteins such as ovalbumin or serum albumin prior to their administration. Therapeutic formulations may be administered in many conventional dosage formulations. While it is possible for the active ingredient to be administered alone, it is preferable to present it as a pharmaceutical formulation. Formulations typically comprise at least one active ingredient, as defined above, together with one or more acceptable carriers thereof. Each carrier should be both pharmaceutically and physiologically acceptable in the sense of being compatible with the other ingredients and not injurious to the patient. Formulations include those suitable for oral, rectal, nasal, topical, or parenteral (including subcutaneous, intramuscular, intravenous and intradermal) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by methods well known in the art of pharmacy. See, e.g., Gilman,

et al. (eds. 1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences, 17th ed. (1990), Mack Publishing Co., Easton, Penn.; Avis, et al. (eds. 1993) Pharmaceutical Dosage Forms: Parenteral Medications, Dekker, New York; Lieberman, et al. (eds. 1990) Pharmaceutical Dosage Forms: Tablets, Dekker, New York; and Lieberman, et al. (eds. 1990) Pharmaceutical Dosage Forms: Disperse Systems, Dekker, New York. The therapy of this invention may be combined with or used in association with other agents, e.g., other modulators of cell activation, e.g., CD40, CD40 ligand, CD28, CTLA-4, B7, B70, SLAM, T cell receptor signaling entities, or their respective antagonists.

Both the naturally occurring and the recombinant forms of the proteins of this invention are particularly useful in kits and assay methods which are capable of screening compounds for binding activity to the proteins. Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds in a short period. See, e.g., Fodor, et al. (1991) Science 251:767-773, which describes means for testing of binding affinity by a plurality of defined polymers synthesized on a solid substrate. The development of suitable assays can be greatly facilitated by the availability of large amounts of purified, soluble proteins or nucleic acids as provided by this invention.

Other methods can be used to determine the critical residues in the substrate, ligand, or receptor binding interactions. Mutational analysis can be performed, e.g., see Somoza, et al. (1993) J. Exp. Med. 178:549-558, to determine specific residues critical in the interaction and/or signaling. This will allow study of both extracellular domains, involved in the soluble ligand interaction, or intracellular domain of a transmembrane form, which provides interactions important in intracellular signaling.

For example, antagonists can normally be found once the antigen has been structurally defined, e.g., by tertiary structure data. Testing of potential interacting analogs is now possible

upon the development of highly automated assay methods using a purified protein. In particular, new agonists and antagonists will be discovered by using screening techniques described herein. Of particular importance are compounds found to have a combined
5 binding affinity for a spectrum of protein molecules, e.g., compounds which can serve as antagonists for species variants of the antigens.

One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with
10 recombinant DNA molecules expressing desired protein. Cells may be isolated which express a selected protein in isolation from other molecules. Such cells, either in viable or fixed form, can be used for standard binding partner binding assays. See also, Parcé, et al. (1989) Science 246:243-247; and Owicki, et al.
15 (1990) Proc. Nat'l Acad. Sci. USA 87:4007-4011, which describe sensitive methods to detect cellular responses.

Another technique for drug screening involves an approach which provides high throughput screening for compounds having suitable binding affinity to a desired target protein, and is
20 described in detail in Geysen, European Patent Application 84/03564, published on September 13, 1984. First, large numbers of different small peptide test compounds are synthesized on a solid substrate, e.g., plastic pins or some other appropriate surface, see Fodor, et al. (1991). Then the pins are reacted with
25 solubilized, unpurified or solubilized, purified target protein, and washed. The next step involves detecting bound protein.

Rational drug design may also be based upon structural studies of the molecular shapes of the protein and other effectors or analogs. Effectors may be other proteins which mediate other
30 functions in response to binding, or other proteins which normally interact. One means for determining which sites interact with specific other proteins is a physical structure determination, e.g., x-ray crystallography or 2 dimensional NMR techniques. These will provide guidance as to which amino acid residues form
35 molecular contact regions. For a detailed description of protein structural determination, see, e.g., Blundell and Johnson (1976) Protein Crystallography, Academic Press, New York.

IX. Kits

This invention also contemplates use of the proteins, fragments thereof, peptides, and their fusion products in a variety of diagnostic kits and methods for detecting, e.g., the presence of protein or binding partner. Typically the kit will have a compartment containing either a described polypeptide or gene segment or a reagent which recognizes one or the other, e.g., fragments or antibodies. Alternatively, kits may be nucleic acid based.

A kit for determining the binding affinity of a test compound to, e.g., an HDTEA84, would typically comprise a test compound; a labeled compound, for example a binding partner or antibody having known binding affinity for HDTEA84; a source of HDTEA84 (naturally occurring or recombinant); and a means for separating bound from free labeled compound, such as a solid phase for immobilizing the molecule. Once compounds are screened, those having suitable binding affinity to the antigen can be evaluated in suitable biological assays, as are well known in the art, to determine whether they act as agonists or antagonists to the HDTEA84 signaling pathway. The availability of recombinant HDTEA84 polypeptides also provide well defined standards for calibrating such assays.

A preferred kit for determining the concentration of, e.g., an HDTEA84 in a sample would typically comprise a labeled compound, e.g., binding partner or antibody, having known binding affinity for the antigen, a source of antigen (naturally occurring or recombinant) and a means for separating the bound from free labeled compound, e.g., a solid phase for immobilizing the HDTEA84. Compartments containing reagents, and instructions, will normally be provided.

Antibodies, including antigen binding fragments, specific for, e.g., the HDTEA84 or fragments, are useful in diagnostic applications to detect the presence of elevated levels of HDTEA84 and/or its fragments. Such diagnostic assays can employ lysates, live cells, fixed cells, immunofluorescence, cell cultures, body fluids, and further can involve the detection of antigens related

to the antigen in serum, or the like. Diagnostic assays may be homogeneous (without a separation step between free reagent and antigen-binding partner complex) or heterogeneous (with a separation step). Various commercial assays exist, such as
5 radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA), enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA), and the like. See, e.g., Van Vunakis, et al. (1980) Meth Enzymol. 70:1-525; Harlow and Lane (1980) Antibodies: A Laboratory Manual, CSH
10 Press, NY; and Coligan, et al. (eds. 1993) Current Protocols in Immunology, Greene and Wiley, NY.

Anti-idiotypic antibodies may have similar use to diagnose presence of antibodies against a described protein, as such may be diagnostic of various abnormal states. Overproduction of
15 prostaglandin transporter may reflect various medical conditions, which may be diagnostic of abnormal physiological states, particularly in proliferative cell conditions such as cancer or abnormal differentiation. For example, leukemias and lymphomas may exhibit altered transporter expression, which may reflect
20 their altered physiology and may provide means to selectively target. Alternatively, overproduction of HDTEA84, HSLJD37R, RANKL, HCC5, MD-1, or MD-2 may result in production of various immunological reactions which may be diagnostic of abnormal physiological states, particularly in proliferative cell
25 conditions such as cancer or abnormal activation or differentiation. Expression levels of DC-PGT, Dubs, or cyclin E2 may likewise be diagnostic of specific therapeutic conditions, advantageous or disadvantageous.

Frequently, the reagents for diagnostic assays are supplied
30 in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled antibody or binding partner, or labeled HDTEA84 is provided. This is usually in conjunction with other additives, such as buffers, stabilizers,
35 materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also contain instructions for proper use and disposal of the contents after

use. Typically the kit has compartments for each useful reagent. Desirably, the reagents are provided as a dry lyophilized powder, where the reagents may be reconstituted in an aqueous medium providing appropriate concentrations of reagents for performing the assay.

Many of the aforementioned constituents of the drug screening and the diagnostic assays may be used without modification or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In these assays, the binding partner, test compound, HDTEA84, or antibodies thereto can be labeled either directly or indirectly.

Possibilities for direct labeling include label groups: radiolabels such as ^{125}I , enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating the bound from the free polypeptide, or alternatively the bound from the free test compound. The polypeptide can be immobilized on various matrixes followed by washing. Suitable matrices include plastic such as an ELISA plate, filters, and beads. See, e.g., Coligan, et al. (eds. 1993) Current Protocols in Immunology, Vol. 1, Chapter 2, Greene and Wiley, NY. Other suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in Rattle, et al. (1984) Clin. Chem. 30:1457-1461, and the double antibody magnetic particle separation as described in U.S. Pat. No. 4,659,678.

Methods for linking proteins or their fragments to the various labels have been extensively reported in the literature and do not require detailed discussion here. Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of thioethers by reaction of a mercapto group

with an activated halogen such as chloroacetyl, or an activated olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Another diagnostic aspect of this invention involves use of
5 oligonucleotide or polynucleotide sequences taken from the
sequence of a described protein. These sequences can be used as
probes for detecting levels of the message in samples from
patients suspected of having an abnormal condition, e.g., cancer
or developmental problem. Since the antigen is a marker for
10 activation, it may be useful to determine the numbers of activated
T cells to determine, e.g., when additional suppression may be
called for. The preparation of both RNA and DNA nucleotide
sequences, the labeling of the sequences, and the preferred size
of the sequences has received ample description and discussion in
15 the literature. See, e.g., Langer-Safer, et al. (1982) Proc.
Nat'l. Acad. Sci. 79:4381-4385; Caskey (1987) Science 236:962-967;
and Wilchek, et al. (1988) Anal. Biochem. 171:1-32.

Alternatively, antibodies may be employed which can recognize
specific duplexes, including DNA duplexes, RNA duplexes, DNA-RNA
20 hybrid duplexes, or DNA-protein duplexes. The antibodies in turn
may be labeled and the assay carried out where the duplex is bound
to a surface, so that upon the formation of duplex on the surface,
the presence of antibody bound to the duplex can be detected. The
use of probes to the novel anti-sense RNA may be carried out in
25 any conventional techniques such as nucleic acid hybridization,
plus and minus screening, recombinational probing, hybrid released
translation (HRT), and hybrid arrested translation (HART). This
also includes amplification techniques such as polymerase chain
reaction (PCR).

30 Diagnostic kits which also test for the qualitative or
quantitative presence of other markers are also contemplated.
Diagnosis or prognosis may depend on the combination of multiple
indications used as markers. Thus, kits may test for combinations
of markers. See, e.g., Viallet, et al. (1989) Progress in Growth
35 Factor Res. 1:89-97. Other kits may be used to evaluate T cell
subsets.

X. Methods for Isolating Substrates/Specific Partners

The DC-PGT should interact with its substrate target. The substrate will be similar to the organic molecules which are subject to transport. The Dubs and cyclin E2 will also be
5 screened for substrate identification.

The HDTEA84, HSLJD37R, and RANKL protein should interact with a TNF ligand, based, e.g., upon its similarity in structure and function to other cell surface antigens exhibiting similar structure and cell type specificity of expression. The MD-1 and
10 MD-2 antigens are related to known proteins, which interact with B cell antigens. Methods to isolate a ligand are made available by the ability to make purified protein for screening programs. Similar techniques will be applicable to the HCC5 chemokine, and the MD-1 and MD-2 surface receptors.

15 Sequences provided herein will allow for screening or isolation of specific ligands. Many methods exist for expression cloning, panning, affinity isolation, or other means to identify a ligand. A two-hybrid selection system may also be applied making appropriate constructs with the available sequences, as
20 appropriate. See, e.g., Fields and Song (1989) Nature 340:245-246.

The broad scope of this invention is best understood with reference to the following examples, which are not intended to limit the invention to specific embodiments.

EXAMPLES

General Methods

- Some of the standard methods are described or referenced, e.g., in Maniatis, et al. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press; Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.), vols. 1-3, CSH Press, NY; Ausubel, et al., Biology, Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (1987 and Supplements) Current Protocols in Molecular Biology, Greene and Wiley, New York; Innis, et al. (eds. 1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, N.Y. Methods for protein purification include such methods as ammonium sulfate precipitation, column chromatography, electrophoresis, centrifugation, crystallization, and others. See, e.g., Ausubel, et al. (1987 and periodic supplements); Deutscher (1990) "Guide to Protein Purification" in Methods in Enzymol. vol. 182, and other volumes in this series; and manufacturer's literature on use of protein purification products, e.g., Pharmacia, Piscataway, N.J., or Bio-Rad, Richmond, CA. Combination with recombinant techniques allow fusion to appropriate segments, e.g., to a FLAG sequence or an equivalent which can be fused via a protease-removable sequence. See, e.g., Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) Genetic Engineering, Principle and Methods 12:87-98, Plenum Press, N.Y.; and Crowe, et al. (1992) QIAexpress: The High Level Expression & Protein Purification System QIAGEN, Inc., Chatsworth, CA. Cell culture techniques are described in Doyle, et al. (eds. 1994) Cell and Tissue Culture: Laboratory Procedures, John Wiley and Sons, NY.
- Standard immunological techniques are described, e.g., in Hertenberg, et al. (eds. 1996) Weir's Handbook of Experimental Immunology vols. 1-4, Blackwell Science; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Methods in Enzymology volumes. 70, 73, 74, 84, 92, 93, 108, 116, 121, 132, 150, 162, and 163.

FACS analyses are described in Melamed, et al. (1990) Flow Cytometry and Sorting Wiley-Liss, Inc., New York, NY; Shapiro (1988) Practical Flow Cytometry Liss, New York, NY; and Robinson, et al. (1993) Handbook of Flow Cytometry Methods Wiley-Liss, New York, NY. Fluorescent labeling of appropriate reagents was performed by standard methods.

The FASTA (Pearson and Lipman, 1988) and BLAST (Altschul, et al. (1990) J. Mol. Biol. 215:403-410) programs were used to comb nonredundant protein and nucleotide databases (Benson, et al. (1994) Nucl. Acids Res. 22:3441-3444; Bairoch and Boeckmann (1994) Nucl. Acids Res. 22:3578-3580) with the resultant cDNA and encoded protein sequences. The sensitive search strategies of Altschul, et al. (1994) Nature Genet. 6:119-129; and Koonin, et al. (1994) EMBO J. 13:493-503; served as examples of how to locate distant structural homologues of protein chains. Multiple alignments of collected homologues were carried out with ClustalW (Thompson, et al. (1994) Comp. Applic. Biosci. 10:19-29) and MACAW (Schuler, et al. (1991) Proteins 9:180-190).

The membrane topologies of proteins, e.g., DC-PGT, and a cohort of putative homologues were analyzed by a variety of methods that sought to determine the consensus number of domains, e.g., hydrophobic membrane-spanning helices and the likely cytoplasmic or extracellular exposure of the hydrophilic connecting loops. For single sequence analysis, the ALOM and MTOP (Klein, et al. (1985) Biochim. Biophys. Acta 815:468-476; and Hartmann, et al. (1989) Proc. Natl. Acad. Sci. USA 86:5786-5790) programs were accessed from the PSORT World-Wide Web site (Nakai and Kanehisa (1991) Proteins 11:95-110; and Nakai and Kanehisa (1992) Genomics 14:897-911); in turn, the TopPredII program (Claros and von Heijne (1994) Comp. Applic. Biosci. 10:685-686; MacIntosh PPC version) was used to parse chains into probable hydrophobic transmembrane and loop regions of DC-PGT, and further predict the localization of these latter regions by prevalence of charged residue types (von Heijne (1992) J. Mol. Biol. 225:487-494; and Sippos and von Heijne (1993) Eur. J. Biochem. 213:1333-

1340) . . MEMSAT (Jones, et al. (1994) Biochem. 33:3038-3049; MS-DOS
PC version) was likewise used to fit individual sequences into
statistically-based topology models that render judgment on
membrane spanning and loop chain segments. Two Web-accessible
5 programs that are able to make use of evolutionary data by
analyzing multiply aligned sequences are PHD (Rost, et al. (1994)
Comp. Applic. Biosci. 10:53-60; and Rost, et al. (1995) Protein
Sci. 4:521-533) and TMAP (Persson and Argos (1994) J. Mol. Biol.
237:182-192); the former utilizes a neural network system to
10 accurately predict the shared location of helical transmembrane
segments in a protein family. Similar analysis of other proteins
can be performed.

I. Generation of Dendritic Cells

15 Human CD34+ cells are obtained as follows. See, e.g., Caux,
et al. (1995) pages 1-5 in Banchereau and Schmitt Dendritic Cells
in Fundamental and Clinical Immunology Plenum Press, NY.
Peripheral or cord blood cells, sometimes CD34+ selected, are
cultured in the presence of Stem Cell Factor (SCF), GM-CSF, and
20 TNF- α in endotoxin free RPMI 1640 medium (GIBCO, Grand Island, NY)
supplemented with 10% (v/v) heat-inactivated fetal bovine serum
(FBS; Flow Laboratories, Irvine, CA), 10 mM HEPES, 2 mM L-
glutamine, 5 X 10⁻⁵ M 2-mercaptoethanol, penicillin (100 μ g/ml).
This is referred to as complete medium.

25 CD34+ cells are seeded for expansion in 25 to 75 cm² flasks
(Corning, NY) at 2 x 10⁴ cells/ml. Optimal conditions are
maintained by splitting these cultures at day 5 and 10 with medium
containing fresh GM-CSF and TNF- α (cell concentration: 1-3 x 10⁵
cells/ml). In certain cases, cells are FACS sorted for CD1a
30 expression at about day 6.

In certain situations, cells are routinely collected after 12
days of culture, eventually adherent cells are recovered using a 5
mM EDTA solution. In other situations, the CD1a+ cells are
activated by resuspension in complete medium at 5 x 10⁶ cells/ml
35 and activated for the appropriate time (e.g., 1 or 6 h) with 1

µg/ml phorbol 12-myristate 13-acetate (PMA, Sigma) and 100 ng/ml ionomycin (Calbiochem, La Jolla, CA). These cells are expanded for another 6 days, and RNA isolated for cDNA library preparation. Other specific cell types may be similarly isolated.

5

II. RNA Isolation and Library Construction

Total RNA is isolated using, e.g., the guanidine thiocyanate/CsCl gradient procedure as described by Chirgwin, et al. (1978) Biochem. 18:5294-5299.

10

Alternatively, poly(A)⁺ RNA is isolated using the OLIGOTEX mRNA isolation kit (QIAGEN). Double stranded cDNA are generated using, e.g., the SUPERScript plasmid system (Gibco BRL, Gaithersburg, MD) for cDNA synthesis and plasmid cloning. The resulting double stranded cDNA is unidirectionally cloned, e.g.,

15

into pSport1 and transfected by electroporation into ELECTROMAX DH10BTM Cells (Gibco BRL, Gaithersburg, MD).

III. Sequencing

DNA isolated from randomly picked clones, or after subtractive hybridization using inactivated cells, are subjected to nucleotide sequence analysis using standard techniques. Alternatively, selected isolated clones can be selected. A Taq DiDeoxy Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) can be used. The labeled DNA fragments are separated using a DNA sequencing gel of an appropriate automated sequencer. Alternatively, the isolated clone is sequenced as described, e.g., in Maniatis, et al. (Current ed.) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press; Sambrook, et al. (Current ed.) Molecular Cloning: A Laboratory Manual, (2d ed.), vols. 1-3, CSH Press, NY; Ausubel, et al., Biology, Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (Current ed., and Supplements) Current Protocols in Molecular Biology, Greene/Wiley, New York. Chemical sequencing methods are also available, e.g., using Maxim and

35

Gilbert sequencing techniques.

IV. Recombinant gene constructs

Poly(A)⁺ RNA is isolated from appropriate cell populations, e.g., using the FastTrack mRNA kit (Invitrogen, San Diego, CA). Samples are electrophoresed, e.g., in a 1% agarose gel containing
5 formaldehyde and transferred to a GeneScreen membrane (NEN Research Products, Boston, MA). Hybridization is performed, e.g., at 65° C in 0.5 M NaHPO₄ pH 7.2, 7% SDS, 1 mM EDTA, and 1% BSA (fraction V) with ³²P-dCTP labeled DC gene cDNA at 10⁷ cpm/ml. After hybridization, filters are washed three times at 50° C in
10 0.2X SSC, 0.1% SDS, e.g., for 30 min, and exposed to film for 24 h. A positive signal will typically be 2X over background, preferably 5-25X.

The recombinant gene construct may be used to generate a probe for detecting the message. The insert may be excised and
15 used in the detection methods described above. Various standard methods for cross species hybridization and washes are well known in the art. See, e.g., Sambrook, et al. and Ausubel.

V: Gene Cloning

20 The HDTEA84 was assembled by careful analysis of ESTs found in various databases. These ESTs were from cDNA libraries derived from Hodgkin's lymphoma, endothelial cells, keratinocytes, prostate, and cerebellum. PCR primers are designed and synthesized and a PCR product is obtained from any of these
25 libraries. This product is used as a hybridization clone to screen these libraries for a full length clone, which may include a transmembrane segment.

Likewise, the HSLJD37R was identified from sequences derived from cDNA libraries from: smooth muscle, pancreas tumor,
30 adipocytes, HUVEC cells, adult pulmonary, endothelial cells, prostate cell line PC3, microvascular endothelial cells, fetal heart, and dendritic cells. A GenBank report by Pan, et al. has been submitted. See GenBank Accession 3549263. Other sequences were detected in libraries from: multiple sclerosis lesions,
35 breast, kidney, and germinal center B cells. RT-PCT showed signal in B cells, PBL, granulocytes, T cells, monocytes, dendritic cell subpopulations including PMA/ionomycin treated, U937 cells, JY

cells, MRC5 cells, CHA, Jurkat, and YC1 cells. This suggests that the transcript is widely expressed.

RANKL was also identified in cDNA libraries from specific tissues, as described. Likewise, the HCC5 chemokine sequence was identified. The Dub11 and Dub12 genes were identified, in part, from their similarity to known Dub1 and Dub2 genes. The MD-1 and MD-2 were identified, in part, from their similarity to the ligand for the RP105 gene. The cyclin E2 was identified based upon its similarity to cyclin E.

VI. Expression Profile

To examine DC-PGT mRNA expression standard Northern Blot Analysis using a RT-PCR fragment of DC-PGT were carried out against human tissue, e.g., Northern blots containing approximately 10 to 20 µg of total RNA are run in formaldehyde gels and transferred to Nytran membranes (Schleicher & Schuell, Keene, NH) by standard methods, and blots were hybridized with a labeled PCR fragment of DC-PGT and washed at 65° C. cDNA can be isolated from cells, embryonic tissues, and adult organs using RNazol solution (Tel-test, Inc., Friendswood, TX) according to manufacturer's instructions. Large amounts of plasmid DNA containing differential display PCR products are prepared using the QIAGEN Plasmid Maxi Kit (QIAGEN) following the manufacturer's instructions. Plasmid DNA is cut with EcoRI (Boehringer Mannheim) or BstXI (NE Biolabs, Mass.), gel extracted with the QIAEX gel extraction kit (QIAGEN) and random primed with [³²P]dCTP (Amersham) using the Prime-It II kit (Stratagene, La Jolla, CA), all in accordance with manufacturer's instructions. Various primers may be used to quantitate expression of message. Means to block DNA hybridization signal, or RNA isolation, will be applicable to quantitate roughly the amount of expression of appropriate RNAs.

The results revealed mRNA of one band at approximately 9.0 kB, another band at approximately 3.0 kB, and a 4.4 kB size which is consistent with the size predicted for the SEQ ID NO: 1 nucleic acid. The smaller mRNA product band could be an alternatively spliced form of SEQ ID NO: 1. DC-PGT is highly expressed in both

activated and non-activated dendritic cells (DC), activated monocytes, activated granulocytes and adult lung. No expression was found in T or PBL cells (either activated or non-activated). Minor expression was detected in B cell (both activated and non-activated) and limited expression was detected in the brain. The results of the northern analysis suggests an expression in macrophages, rather than monocytes (Kuppfer cells in the liver, microglial cells in the brain, alveolar macrophages in the lung) particularly as there is no expression in PBL. Southern expression analysis carried out using common techniques confirmed the expression pattern revealed in the Northern analysis.

For example, the DC-PGT tissue distribution seems to have highest mRNA levels in kidney, placenta, liver, bone marrow, thymus, spleen, lung, and some in testis. This distribution corresponds to organs with especially important ion exchange features, e.g., Na, K, or Ca, or in hematopoietic organs. Generally, the expression is higher in fibroblast and hematopoietic cells compared to neuronal cells.

A probe specific for cDNA encoding the HDTEA84, HSLJD37R, or RANKL is used to determine tissue distribution of message encoding the antigen. Standard hybridization probes may be used to do a Northern analysis of RNA from appropriate sources, either cells, e.g., stimulated, or in various physiological states, in various tissues, e.g., spleen, liver, thymus, lung, etc., or in various species. Southern analysis of cDNA libraries may also provide valuable distribution information. Standard tissue blots or species blots are commercially available. Similar techniques will be useful for evaluating diagnostic or medical conditions which may correlate with expression in various cell types.

PCR analysis using appropriate primers may also be used. Antibody analysis, including immunohistochemistry or FACS, may be used to determine cellular or tissue distribution.

Southern blot analysis of primate cDNA libraries is performed on, e.g.,: U937 premonocytic line, resting (M100); elutriated monocytes, activated with LPS, IFN γ , anti-IL-10 for 4, 16 h pooled (M106); elutriated monocytes, activated with LPS, IFN γ , IL-10 for 4, 16 h pooled (M107); elutriated monocytes, activated LPS for 1 h

(M108); elutriated monocytes, activated LPS for 6 h (M109); dendritic cells (DC) 30% CD1a+, from CD34+ GM-CSF, TNF α 12 days, resting; DC 70% CD1a+, from CD34+ GM-CSF, TNF α 12 days, resting (D101); DC 70% CD1a+, from CD34+ GM-CSF, TNF α 12 days, activated with PMA and ionomycin for 1 hr (D102); DC 70% CD1a+, from CD34+ GM-CSF, TNF α 12 days, activated with PMA and ionomycin for 6 hr (D103); DC 95% CD1a+, from CD34+ GM-CSF, TNF α 12 days activated with PMA and ionomycin for 1 or 6 hr, pooled; DC from monocytes GM-CSF, IL-4 5 days, resting (D107); DC from monocytes GM-CSF, IL-4 5 days, resting (D108); DC from monocytes GM-CSF, IL-4 5 days, activated TNF α , monocyte supe for 4, 16 h pooled (D110); EBV transfected B cell lines, resting; spleenocytes, resting; spleenocytes, activated with PMA and ionomycin; 20 NK clones resting, pooled; 20 NK clones activated with PMA and ionomycin, pooled; NKL clone, IL-2 treated; NK cytotoxic clone, resting; adipose tissue fetal 28 wk male (O108); brain fetal 28 wk male (O104); gallbladder fetal 28 wk male (O106); heart fetal 28 wk male (O103); small intestine fetal 28 wk male (O107); kidney fetal 28 wk male (O100); liver fetal 28 wk male (O102); lung fetal 28 wk male (O101); ovary fetal 25 wk female (O109); adult placenta 28 wk (O113); spleen fetal 28 wk male (O112); testes fetal 28 wk male (O111); uterus fetal 25 wk female (O110); TH0 clone Mot 72, resting (T102); T cell, TH0 clone Mot 72, activated with anti-CD28 and anti-CD3 for 3, 6, 12 h pooled (T103); T cell, TH0 clone Mot 72, anergic treated with specific peptide for 2, 7, 12 h pooled (T104); Th0 subtraction of resting from activated; T cell, TH1 clone HY06, resting (T107); T cell, TH1 clone HY06, activated with anti-CD28 and anti-CD3 for 3, 6, 12 h pooled (T108); T cell, TH1 clone HY06, anergic treated with specific peptide for 2, 6, 12 h pooled (T109); Th1 subtraction of resting from activated; T cell, TH2 clone HY935, resting (T110); T cell, TH2 clone HY935, activated with anti-CD28 and anti-CD3 for 2, 7, 12 h pooled (T111); and Th2 subtraction of resting from activated.

Samples for mouse mRNA distribution may include, e.g.,:
 35 resting mouse fibroblastic L cell line (C200); Braf:ER (Braf

fusion to estrogen receptor) transfected cells, control (C201); T cells, TH1 polarized (Mel14 bright, CD4+ cells from spleen, polarized for 7 days with IFN- γ and anti IL-4; T200); T cells, TH2 polarized (Mel14 bright, CD4+ cells from spleen, polarized for 7 days with IL-4 and anti-IFN- γ ; T201); T cells, highly TH1 polarized (see Openshaw, et al. (1995) J. Exp. Med. 182:1357-1367; activated with anti-CD3 for 2, 6, 16 h pooled; T202); T cells, highly TH2 polarized (see Openshaw, et al. (1995) J. Exp. Med. 182:1357-1367; activated with anti-CD3 for 2, 6, 16 h pooled; T203); CD44- CD25+ pre T cells, sorted from thymus (T204); TH1 T cell clone D1.1, resting for 3 weeks after last stimulation with antigen (T205); TH1 T cell clone D1.1, 10 μ g/ml ConA stimulated 15 h (T206); TH2 T cell clone CDC35, resting for 3 weeks after last stimulation with antigen (T207); TH2 T cell clone CDC35, 10 μ g/ml ConA stimulated 15 h (T208); Mel 14+ naive T cells from spleen, resting (T209); Mel14+ T cells, polarized to Th1 with IFN- γ /IL-12/anti-IL-4 for 6, 12, 24 h pooled (T210); Mel 14+ T cells, polarized to Th2 with IL-4/anti-IFN- γ for 6, 13, 24 h pooled (T211); unstimulated mature B cell leukemia cell line A20 (B200); unstimulated B cell line CH12 (B201); unstimulated large B cells from spleen (B202); B cells from total spleen, LPS activated (B203); metrizamide enriched dendritic cells from spleen, resting (D200); dendritic cells from bone marrow, resting (D201); monocyte cell line RAW 264.7 activated with LPS 4 h (M200); bone-marrow macrophages derived with GM and M-CSF (M201); macrophage cell line J774, resting (M202); macrophage cell line J774 + LPS + anti-IL-10 at 0.5, 1, 3, 6, 12 h pooled (M203); macrophage cell line J774 + LPS + IL-10 at 0.5, 1, 3, 5, 12 h pooled (M204); aerosol challenged mouse lung tissue, Th2 primers, aerosol OVA challenge 7, 14, 23 h pooled (see Garlisi, et al. (1995) Clinical Immunology and Immunopathology 75:75-83; X206); Nippostrongylus-infected lung tissue (see Coffman, et al. (1989) Science 245:308-310; X200); total adult lung, normal (O200); total lung, rag-1 (see Schwarz, et al. (1993) Immunodeficiency 4:249-252; O205); IL-10 K.O. spleen (see Kuhn, et

al. (1991) Cell 75:263-274; X201); total adult spleen, normal (O201); total spleen, rag-1 (O207); IL-10 K.O. Peyer's patches (O202); total Peyer's patches, normal (O210); IL-10 K.O. mesenteric lymph nodes (X203); total mesenteric lymph nodes, normal (O211); IL-10 K.O. colon (X203); total colon, normal (O212); NOD mouse pancreas (see Makino, et al. (1980) Jikken Dobutsu 29:1-13; X205); total thymus, rag-1 (O208); total kidney, rag-1 (O209); total heart, rag-1 (O202); total brain, rag-1 (O203); total testes, rag-1 (O204); total liver, rag-1 (O206); rat normal joint tissue (O300); and rat arthritic joint tissue (X300).

A. Direct protein detection by antibodies

Various cells, tissues, and developmental stages are stained with labeled antibodies. The detection may be immuno-histochemical for solid tissue, by FACS in disperse cells, and by other appropriate methods for other sample types. Antibodies specific for the various forms may be used to distinguish between membrane associated and soluble fragments. Various amplification means may be coupled to increase sensitivity.

B. Functional detection

Specific neutralizing antibodies should provide means to specifically block the biological activity of the prostaglandin transporter. Activities related to prostaglandin binding, or to prostaglandin transport may be measured by sensitive means based upon knowledge of the normal biological function of the various forms.

Further testing of populations of cells, e.g., hematopoietic progenitors, or of other cell or tissue types will be useful to further determine distribution and likely function. Other tissue types, at defined developmental stages, and pathology samples may be screened to determine whether pathological states or stages may be advantageously correlated with the biological activity of the transporter.

VII. Protein Expression

PCR is used to make a construct comprising the open reading frame, preferably in operable association with proper promoter, selection, and regulatory sequences. The resulting expression

plasmid is transformed into an appropriate cell type, e.g., the Topp5, E. coli strain (Stratagene, La Jolla, CA). Ampicillin resistant (50 µg/ml) transformants are grown in Luria Broth (Gibco) at 37° C until the optical density at 550 nm is 0.7.

5 Recombinant protein is induced with 0.4 mM

isopropyl-βD-thiogalacto-pyranoside (Sigma, St. Louis, MO) and incubation of the cells continued at 20° C for a further 18 hours. Cells from a 1 liter culture are harvested by centrifugation and resuspended, e.g., in 200 ml of ice cold 30% sucrose, 50 mM Tris HCl pH 8.0, 1 mM ethylenediaminetetraacetic acid. After 10 min on ice, ice cold water is added to a total volume of 2 liters. After 20 min on ice, cells are removed by centrifugation and the supernatant is clarified by filtration via a 5 µM Millipak 60 (Millipore Corp., Bedford, MA).

15 The recombinant protein is purified via standard purification methods, e.g., various ion exchange chromatography methods. Immunoaffinity methods using antibodies described below can also be used. Affinity methods may be used where an epitope tag is engineered into an expression construct.

20 Similar methods are used to prepare expression constructs and cells in eukaryotic cells. Eukaryotic promoters and expression vectors may be produced, as described above.

Further study of the expression and control of prostaglandin transporter will be pursued. The controlling elements associated with the antigens may exhibit differential developmental, tissue specific, or other expression patterns. Upstream or downstream genetic regions, e.g., control elements, are of interest.

Multiple transfected cell lines are screened for one which expresses the antigen, membrane bound, or soluble forms, at a high level compared with other cells. Various cell lines are screened and selected for their favorable properties in handling. Natural protein can be isolated from natural sources, or by expression from a transformed cell using an appropriate expression vector. Purification of the expressed protein is achieved by standard procedures, or may be combined with engineered means for effective purification at high efficiency from cell lysates or supernatants. FLAG or His₆ segments can be used for such purification features.

VIII. Protein Purification

The prostaglandin transporter is isolated by a combination of affinity chromatography using the prostaglandin transporter specific binding compositions, e.g., antibody, as a specific binding reagent in combination with protein purification techniques allowing separation from other proteins and contaminants. Various detergent combinations may be tested to determine what combinations will retain biological activity while solubilizing contaminants. The purification may follow biological activity, e.g., prostaglandin binding or transport into membranes, or by ELISA or other structural binding reagents.

Similar methods are applied for purification of other polypeptides.

IX. Isolation of Homologous Genes

The described genes, e.g., cDNA, can be used as a hybridization probe to screen a library from a desired source, e.g., a primate cell cDNA library. Many different species can be screened both for stringency necessary for easy hybridization, and for presence using a probe. Appropriate hybridization conditions will be used to select for clones exhibiting specificity of cross hybridization.

cDNA libraries from the desired species are collected, from appropriate cell types. Screening by hybridization or PCR using degenerate probes based upon the peptide sequences will also allow isolation of appropriate clones. Alternatively, use of appropriate primers for PCR screening will yield enrichment of appropriate nucleic acid clones.

Similar methods are applicable to isolate either species, polymorphic, or allelic variants. Species variants are isolated using cross-species hybridization techniques based upon isolation of a full length isolate or fragment from one species as a probe.

Alternatively, antibodies raised against proteins will be used to screen for cells which express cross-reactive proteins from an appropriate, e.g., cDNA library. The purified protein or defined peptides are useful for generating antibodies by standard

methods, as described above. Synthetic peptides or purified protein are presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press. The resulting antibodies are used, e.g., for screening, panning, or sorting.

X. Antibody Preparation

Synthetic peptides or purified protein, natural or recombinant, are presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press. Polyclonal serum, or hybridomas may be prepared. In appropriate situations, the binding reagent is either labeled as described above, e.g., fluorescence or otherwise, or immobilized to a substrate for panning methods.

XI. Chromosome Mapping

DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer and hybridization are performed according to standard techniques. See Jenkins, et al. (1982) J. Virol. 43:26-36. Blots may be prepared with Hybond-N nylon membrane (Amersham). The probe is labeled with ^{32}P -dCTP; washing is done to a final stringency, e.g., of 0.1X SSC, 0.1% SDS, 65° C.

Alternatively, a BIOS Laboratories (New Haven, CT) mouse somatic cell hybrid panel may be combined with PCR methods. See Fan, et al. (1996) Immunogenetics 44:97-103.

Chromosome spreads are prepared. In situ hybridization is performed on chromosome preparations obtained from phytohemagglutinin-stimulated human lymphocytes cultured for 72 h. 5-bromodeoxyuridine is added for the final seven hours of culture (60 $\mu\text{g/ml}$ of medium), to ensure a posthybridization chromosomal banding of good quality.

A PCR fragment, amplified with the help of primers, is cloned into an appropriate vector. The vector is labeled by nick-translation with ^3H . The radiolabeled probe is hybridized to metaphase spreads at final concentration of 200 ng/ml of hybridization solution as described in Mattei, et al. (1985) Hum. Genet. 69:327-331.

After coating with nuclear track emulsion (KODAK NTB₂), slides are exposed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads are first stained with buffered Giemsa solution and metaphase photographed. R-banding is then performed by the fluorochrome-photolysis-Giemsa (FPG) method and metaphases rephotographed before analysis.

Using these techniques, the DC-PGT gene was mapped to marker SHGC-3911 on chromosome 11q13 with a resulting lod score of 1000.0. Other markers in the SHGC-3911 region at chromosome 11q13 include the FcεRI receptor which is alleged to be associated with allergic conditions. In comparison to the location of DC-PGT, the ubiquitously expressed human PGT homologue of Lu et al., (described above) is localized to chromosome 7.

XII. Biochemical Characterization

Constructs for the expression of, e.g., DC-PGT are made with a tag (FLAG) sequence (Hopp, et al. (1988) Biotechnology (NY) 6:1205-1210) introduced in the protein. The open reading frame of the DC-PGT cDNA of SEQ ID NO: 1 is amplified by appropriate PCR primers using standard methods to introduce the FLAG peptide sequence (IBI, New Haven, CT) at the C-terminus of the protein. For example, a PFU enzyme (Stratagene) with 12 cycles PCR: 94° C 30 sec; 55° C 1 min; 72° C 4 min. PCR constructs are cloned into a PME18X vector (DNAX) using XhoI and XbaI sites incorporated into the 5' and 3' primers, respectively.

COS-7 cells are maintained in DMEM, 10% FCS, 4 mM L-glutamine (JRH Biosciences, Lenexa, KS), 100 U/ml penicillin, and 100 µg/ml streptomycin. Plasmid DNA is transfected by electroporation (BIORAD, Hercules, CA) (20 µg / 1×10^7 cells) and plated into tissue culture dishes. The medium is replaced after 24 hours and

cell lysates and media are collected three days after transfection. Lysis buffer (25 mM Hepes pH 7.5, 2 mM EDTA, 1.0% NP-40, 150 mM NaCl, 0.01% Aprotinin (Sigma, St. Louis, MO), 0.01% Leupeptin (Sigma)) is added to the plates. Plates are kept on ice for 45 minutes. Lysates are centrifuged for 15 minutes to eliminate cell debris. Supernatants of centrifuged cell lysates and sterile-filtered media from cultured cells are incubated with anti-FLAG M2 Affinity Gel (IBI) at 4° C overnight and washed four times with PBS. Immunoprecipitates are eluted in a Econocolumn (BIORAD) with 2.5 M Glycine, pH 2.5. Eluates are neutralized with Hepes, pH 7.4 (JRH Biosciences) and concentrated by precipitation with 24% TCA and 2% deoxycholic sodium salt (Sigma). Pellets are eluted in 2 x Sample Buffer (NOVEX, San Diego, CA), electrophoresed on 4-20% tris-glycine gels (Novex) and transferred to PVDF membranes (Immobilon-P, Millipore Corporation, Bedford, MA). Membranes are exposed to 3% non-fat milk for 1 h at 37° C. Anti-FLAG M2 antibody is used as recommended (IBI). Anti-mouse Ig horseradish peroxidase conjugate (Amersham) is used at 1:2,000 dilution and the peroxidase detection is performed with ECL detection reagents (Amersham).

Other fusion proteins can be produced, e.g., a recombinant prostaglandin transporter construct is prepared, e.g., as a fusion product with a useful affinity reagent, e.g., FLAG peptide. This peptide segment may be useful for purifying the expression product of the construct. See, e.g., Crowe, et al. (1992) QIAexpress: The High Level Expression & Protein Purification System QUIAGEN, Inc. Chatsworth, CA; and Hopp, et al. (1988) Bio/Technology 6:1204-1210. Membranes comprising the transporter are assayed to determine the natural prostaglandin substrate. Most likely the prostaglandin will be a uracil related prostaglandin, but may also include, at various levels of efficiency of binding or transport, pyrimidine or purine analogs. See, e.g., Goodman and Gilman (Current ed.), The Pharmacological Basis of Therapeutics; Lukovics and Zablocka Nucleoside Synthesis: Organosilicon Methods Ellis Horwood, N.Y.; Townsend, Chemistry of Nucleosides and Nucleotides, vols. 1-3, Plenum Press, N.Y.; Munch-Pertson (1983) Metabolism of Nucleotides, Nucleosides, and Prostaglandins in Microorganisms

Academic Press, NY; Gehrke (1990) Chromatography & Modification of Nucleosides vols. A, B, and C, Elsevier; Bloch (1975) Chemistry, Biology, & Clinical Uses of Nucleoside Analogs Annals NY Acad. Sci.; and Ulbricht (1964) Purines, Pyrimidines, & Nucleotides
5 Franklin Co.

XIII. Expression Cloning; Partner Screening

A. Antibodies and flow-cytometric sorting

Expression cloning of cells transformed with an appropriate
10 cDNA library may be sorted by FACS using antibody reagents described above. The sorted cells are isolated and expanded, and subjected to multiple selection cycles, leading to a high proportion of cells expressing the desired DNA.

B. Antibodies and staining

15 The antibodies to, e.g., DC-PGT, are used for screening of a library made from a cell line which expresses the polypeptide. Standard staining techniques are used to detect or sort intracellular or surface expressed ligand, or surface expressing transformed cells are screened by panning. Screening of
20 intracellular expression is performed by various staining or immunofluorescence procedures. See also McMahan, et al. (1991) EMBO J. 10:2821-2832.

For example, on day 0, precoat 2-chamber permanox slides with 1 ml per chamber of fibronectin, 10 ng/ml in PBS, for 30 min at
25 room temperature. Rinse once with PBS. Then plate COS cells at $2-3 \times 10^5$ cells per chamber in 1.5 ml of growth media. Incubate overnight at 37° C.

On day 1 for each sample, prepare 0.5 ml of a solution of 66 µg/ml DEAE-dextran, 66 µM chloroquine, and 4 µg DNA in serum free
30 DME. For each set, a positive control is prepared, e.g., of huIL-10-FLAG cDNA at 1 and 1/200 dilution, and a negative mock. Rinse cells with serum free DME. Add the DNA solution and incubate 5 hr at 37° C. Remove the medium and add 0.5 ml 10% DMSO in DME for 2.5 min. Remove and wash once with DME. Add 1.5 ml growth medium
35 and incubate overnight.

On day 2, change the medium. On days 3 or 4, the cells are fixed and stained. Rinse the cells twice with Hank's Buffered Saline Solution (HBSS) and fix in 4% paraformaldehyde (PFA)/glucose for 5 min. Wash 3X with HBSS. The slides may be stored at -80° C after all liquid is removed. For each chamber, 0.5 ml incubations are performed as follows. Add HBSS/saponin (0.1%) with 32 µl/ml of 1M NaN₃ for 20 min. Cells are then washed with HBSS/saponin 1X. Soluble antibody is added to cells and incubate for 30 min. Wash cells twice with HBSS/saponin. Add second antibody, e.g., Vector anti-mouse antibody, at 1/200 dilution, and incubate for 30 min. Prepare ELISA solution, e.g., Vector Elite ABC horseradish peroxidase solution, and preincubate for 30 min. Use, e.g., 1 drop of solution A (avidin) and 1 drop solution B (biotin) per 2.5 ml HBSS/saponin. Wash cells twice with HBSS/saponin. Add ABC HRP solution and incubate for 30 min. Wash cells twice with HBSS, second wash for 2 min, which closes cells. Then add Vector diaminobenzoic acid (DAB) for 5 to 10 min. Use 2 drops of buffer plus 4 drops DAB plus 2 drops of H₂O₂ per 5 ml of glass distilled water. Carefully remove chamber and rinse slide in water. Air dry for a few minutes, then add 1 drop of Crystal Mount and a cover slip. Bake for 5 min at 85-90° C.

Alternatively, the antibodies to a selected protein are used to affinity purify or sort out cells expressing the antigen. See, e.g., Sambrook et al. or Ausubel et al, which are incorporated herein by reference. The antigen is typically expressed on the cell surface.

Hybridization approaches may also be utilized to find closely related variants of the antigen based upon nucleic acid hybridization.

XIV. Screening for DC-PGT Substrate Specificity

The types of organic anions transported by DC-PGT of the present invention can be directly tested using standard methods. For example, DC-PGT cDNAs can be expressed in HeLa cell monolayers or in Xenopus oocytes to determine the ability of DC-PGT to uptake various tracer labeled substrates e.g., prostaglandins such as PGE₁, PGE₂, PGE_{2a}, PGD₂, thromboxanes such as Tx_{B2} or non-

prostaglandin anionic substrates such as glutathione, p-amino hippurate, taurochoalate, urate, unconjugated and conjugated bilirubin, and estradiol glucouronide. For example, for oocyte expression, water or complementary RNA (cRNA) that has been

5 transcribed in vitro from DC-PGT cDNA and capped is injected into *Xenopus* oocytes at approximately 50 ng of cRNA per oocyte. Uptake studies are performed 2 to 3 days after injection by washing of oocytes three times in Waymouth's solution, incubating for various periods at 27°C with radioactive substrates (approx. 0.25 μ Ci/ml; total concentration, approx. 1 nM), washing three times with ice-cold Waymouth's solution, and lysing in 0.5 ml of 10% SDS.

10 Oocyte-associated radioactivity is determined by liquid scintillation spectroscopy. For HeLa cell expression, cells are grown to approx. 80% confluence on 35 mm dishes then infected with recombinant vaccinia virus vTF7-3 of 10 plaque forming units per cell according to a method of Fuesst, et al. (1986) Proc. Nat'l Acad. Sci. USA 83:8122-8126. Thirty minutes after infection cells are transfected with DC-PGT cDNA (10 μ g/ml) plus lipofectin (20 μ g/ml) according to a method of Blakely, et al. (1991) Anal.

20 Biochem. 194:302-310. After 3 hours of incubation, vaccinia virus and the DNA-lipofectin complex are removed, and the cells are maintained overnight in Dulbecco's modified Eagle's medium supplemented with 5% fetal bovine serum. Uptake studies are performed 19 hours after transfection. Monolayers are washed

25 three times with culture medium without serum and incubated for various times at 27° C with radioactive substrate (0.5 μ Ci/ml per dish; total concentration, approx. 0.2 nM). Uptake is stopped by washing cells once with ice-cold Waymouth's solution containing 5% bovine serum albumin and then four times with Waymouth's solution

30 alone. Cells are scrapped and the associated radioactivity is measured by liquid scintillation spectroscopy.

XV. Measuring DC-PGT Substrate Uptake Kinetics

Competitive tracer uptake kinetics using DC-PGT comparing

35 various prostaglandins or thromboxanes (e.g., PGE1, PGE2, PGE2a, PGD2 or TxB2) are determined using standard competitive transport

assays. For example for determining time dependent uptake of tracer labeled prostaglandin uptakes into HeLa cells expressing DC-PGT clones the following $^3\text{[H]}$ -PGs final concentrations are used (New England Nuclear, Boston, MA): PGE₂: 0.7 nM (176 cpm/fmol); PGE₁: 0.6 nM (62 cpm/fmol); PGD₂: 0.9 nM (126 cpm/fmol); PGF₂ α : 0.6 nM (185 cpm/fmol); TXB₂: 1.0 nM (114 cpm/fmol); PGI₂ analog $^3\text{[H]}$ -iloprost (Amersham Corp., Arlington Heights, IL) at 7.9 nM (14 cpm/fmol).

10 XVI. Determining DC-PGT uptake inhibition

Compositions inhibiting DC-PGT uptake can also be measured. For example to measure the inhibition of tracer PGE₂, uptakes at 10 min intervals (0.2 nM $^3\text{[H]}$ -PGE₂) with or without various concentrations of unlabeled prostanoids PGE₂, PGE₁, PGD₂, PGF₂ α , TXB₂, PGI₂, (100-500 nM; Cayman Chemical, Ann Arbor MI) or inhibitors such as furosemide, probenecid, and indomethacin (10-100 μM , Sigma Chemical Co., St. Louis, MO) are determined in duplicate on a given transfection for one or two separate transfections. Since the substrate concentrations are at least 500 times less than the concentration of unlabeled prostanoids an apparent affinity constant, $K_{1/2}$ is determined from the equation: $K_{1/2} = [v_i / (v - v_i)] [i]$ where v = uptake without inhibitor, v_i = uptake with inhibitor, and i = inhibitor concentration as described by Neame and Richards (1972) in Elementary Kinetics of Membrane Carrier Transport, John Wiley & Sons, New York.

XVII. Screening for Agonists or Antagonists

Using a HeLa or Xenopus system, described above, or a comparable system, one of ordinary skill in the art can use the DC-PGT of the invention to screen for inhibitors or agonists of DC-PGT mediated tracer transport. The efficacy of potential antagonists can be compared with known PG transport inhibitors such as furosemide, probenecid, or indomethacin. Potential agonist or antagonist compositions are incubated, using a system as described above, for a time sufficient to allow binding of the

test composition and the DC-PGT transporter. Enhancement or decrement in measures of tracer uptake can be correlated to the specific composition being tested. Accordingly, one can identify compounds or compositions that modulate organic anion transport via the DC-PGT transporter of the invention by assessing the uptake of various anions such as prostaglandins or thromboxanes in the presence and absence of the compound or compositions being tested. Similar methods may be used to screen for substrates for the enzymes, e.g., Dubs and cyclin E2.

XVIII. Isolation of Ligand for Receptor

A construct for expression of the product can be used as a specific binding reagent to identify its binding partner, e.g., ligand, by taking advantage of its specificity of binding, much like an antibody would be used. A receptor reagent is either labeled as described above, e.g., fluorescence or otherwise, or immobilized to a substrate for panning methods. See also Anderson, et al. (1997) Nature 390:175-179, which is incorporated herein by reference.

The binding composition is used to screen an expression library made from a cell line which expresses a binding partner, e.g., TNF family ligand. Standard staining techniques are used to detect or sort intracellular or surface expressed receptor, or surface expressing transformed cells are screened by panning. Screening of intracellular expression is performed by various staining or immunofluorescence procedures. See also McMahan, et al. (1991) EMBO J. 10:2821-2832.

Alternatively, receptor reagents are used to affinity purify or sort out cells expressing a receptor. See, e.g., Sambrook, et al. or Ausubel, et al.

Another strategy is to screen for a membrane bound ligand by panning. The cDNA containing ligand cDNA is constructed as described above. The ligand can be immobilized and used to immobilize expressing cells. Immobilization may be achieved by use of appropriate antibodies which recognize, e.g., a FLAG sequence or a receptor fusion construct, or by use of antibodies raised against the first antibodies. Recursive cycles of

selection and amplification lead to enrichment of appropriate clones and eventual isolation of ligand expressing clones.

Phage expression libraries can be screened by receptor. Appropriate label techniques, e.g., anti-FLAG antibodies, will allow specific labeling of appropriate clones.

IX. Chemotaxis Assays

Chemokine proteins are produced, e.g., in COS cells transfected with a plasmid carrying the chemokine cDNA by electroporation. See, Hara, et al. (1992) EMBO J. 10:1875-1884. Physical analytical methods may be applied, e.g., CD analysis, to compare tertiary structure to other chemokines to evaluate whether the protein has likely folded into an active conformation. After transfection, a culture supernatant is collected and subjected to bioassays. A mock control, e.g., a plasmid carrying the luciferase cDNA, is used. See, de Wet, et al. (1987) Mol. Cell. Biol. 7:725-757. A positive control, e.g., recombinant murine MIP-1 α from R&D Systems (Minneapolis, MN), is typically used. Likewise, antibodies may be used to block the biological activities, e.g., as a control.

Lymphocyte migration assays are performed as previously described, e.g., in Bacon, et al. (1988) Br. J. Pharmacol. 95:966-974. Other trafficking assays are also available. See, e.g., Quidling-Järbrink, et al. (1995) Eur. J. Immunol. 25:322-327; Koch, et al. (1994) J. Clinical Investigation 93:921-928; and Antony, et al. (1993) J. Immunol. 151:7216-7223. Murine Th2 T cell clones, CDC-25 (see Tony, et al. (1985) J. Exp. Med. 161:223-241) and HDK-1 (see Cherwinski, et al. (1987) J. Exp. Med. 166:1229-1244), made available from R. Coffman and A. O'Garra (DNAX, Palo Alto, CA), respectively, are used as controls.

Ca²⁺ flux upon chemokine stimulation is measured according to the published procedure described in Bacon, et al. (1995) J. Immunol. 154:3654-3666.

Maximal numbers of migrating cells in response to MIP-1 α typically occur at a concentration of 10⁻⁸ M, in agreement with

original reports for CD4+ populations of human T cells. See Schall (1993) J. Exp. Med. 177:1821-1826. A dose-response curve is determined, preferably giving a characteristic bell shaped dose-response curve.

5 After stimulation with CC chemokines, lymphocytes generally show a measurable intracellular Ca²⁺ flux. MIP-1 α is capable of inducing immediate transients of calcium mobilization. Typically, the levels of chemokine used in these assays will be comparable to those used for the chemotaxis assays (1/1000 dilution of
10 conditioned supernatants).

XX. Biological Activities

A robust and sensitive assay is selected as described above, e.g., on immune cells, neuronal cells, or stem cells. Chemokine
15 is added to the assay in increasing doses to see if a dose response is detected. For example, in a proliferation assay, cells are plated out in plates. Appropriate culture medium is provided, and chemokine is added to the cells in varying amounts. Growth is monitored over a period of time which will detect either
20 a direct effect on the cells, or an indirect effect of the chemokine.

Alternatively, an activation assay or attraction assay is used. An appropriate cell type is selected, e.g., hematopoietic cells, myeloid (macrophages, neutrophils, polymorphonuclear cells,
25 etc.) or lymphoid (T cell, B cell, or NK cells); neural cells (neurons, neuroglia, oligodendrocytes, astrocytes, etc.), or stem cells, e.g., progenitor cells which differentiate to other cell types, e.g., gut crypt cells and undifferentiated cell types.

Retroviral infection assays have also been described using, e.g., the CCR1, CCR3, and CCR5 receptors. These receptors, which
30 bind the RANTES and MIP-1 related chemokines, are likely also to be receptors for the HCC5. Recent description of these chemokine receptors in retroviral infection processes, and the effects by the related RANTES and MIP-1 chemokines, suggest similar effects.
35 may exist with the HCC5. See, e.g., Balter (1996) Science 272:1740 (describing evidence for chemokine receptors as coreceptors for HIV); and Deng, et al. (1996) Nature 381:661-666.

Chemokines may also be assayed for activity in hemopoietic assays as described, e.g., by H. Broxmeyer. See Bellido, et al. (1995) J. Clinical Investigation 95:2886-2895; and Jilka, et al. (1995) Expt'l Hematology 23:500-506. They may be assayed for angiogenic activities as described, e.g., by Streiter, et al. (1992) Am. J. Pathol. 141:1279-1284. Or for a role in inflammation. See, e.g., Wakefield, et al. (1996) J. Surgical Res. 64:26-31.

Other assays will include those which have been demonstrated with other chemokines. See, e.g., Schall and Bacon (1994) Current Opinion in Immunology 6:865-873; and Bacon and Schall (1996) Int. Arch. Allergy & Immunol. 109:97-109.

The DUB genes will be screened for the deubiquitinating activities, as described. See, e.g., Hochstrasser (1995) Curr. Opin. Cell Biol. 7:215-223; Wilkinson, et al. (1995) Biochemistry 34:14535-14546; Baker, et al. (1992) J. Biol. Chem. 267:23364-23375; Baek et al. (1998) J. Biol. Chem. 272:25560-25565; and Papa and Hochstrasser (1993) Nature 366:313-319. For example, for an in vitro assay for UBP Activity, ¹²⁵I-labeled Ub-PESTc is used as a substrate according to the method of Woo, et al. (1995) J. Biol. Chem. 270:18766-18773. Reaction mixtures (0.1 ml) contain the proper amount of the enzyme preparations and 10-30 µg of ¹²⁵I-labeled Ub-PESTc in 100 mM Tris-HCl (pH 7.8), 1 mM dithiothreitol, 1 mM EDTA, and 5% glycerol. After incubating the mixtures for appropriate periods, the reaction is terminated by adding 50 µl of 40% (w/v) trichloroacetic acid and 50 µl of 1.2% (w/v) bovine serum albumin. The samples are centrifuged, and the resulting supernatants are counted for their radioactivities using a counter. The enzyme activity is expressed as a percentage of ¹²⁵I-labeled Ub-PESTc hydrolyzed to acid-soluble products. When assaying the hydrolysis of Ub-NH-carboxyl extension proteins and His-di-Ub, incubations are performed as above but in the presence of 5 µg of the substrate. After incubation for appropriate periods, the samples are subjected to discontinuous gel electrophoresis as described by Baek, et al. (1998) J. Biol. Chem.

272:25560-25565. Proteins in the gels were then visualized by staining with Coomassie Blue R-250 or by exposing to x-ray films (Fuji) at 70° C. To prepare ¹²⁵I-labeled poly-Ub-NH-lysozyme conjugates, 2 µg of the ¹²⁵I-labeled lysozyme (5 x 10⁵ cpm) are
5 incubated with 10 µg of Ub, 120 µg of fraction II, and an ATP-regenerating system consisting of 10 mM Tris-HCl (pH 7.8), 15 units/ml creatine phosphokinase, 6.5 mM phosphocreatine, 1.5 mM ATP, 1 mM dithiothreitol, 0.5 mM MgCl₂, and 1 mM KCl in a final
10 volume of 0.05 ml. Incubations are performed for 2 h at 37° C in the presence of 1 mM hemin to prevent proteolysis of the ubiquitinated protein conjugates by the 26 S proteasome. After incubation, the samples are heated for 10 min at 55° C for inactivation of endogenous UBPs. Alternatively, Dub11 or Dub12 can be expressed as a GST fusion protein according to the method
15 of Zhu, et al. (1997) J. Biol. Chem. 272:51-57 by cloning into an appropriate expression vector and subsequently co-transformed with a plasmid encoding Ub-Met-β-gal, in which ubiquitin is fused to the NH₂ terminus of β-galactosidase and testing for cleavage.

However, the deubiquitinating enzymes have also been reported
20 to have additional functions besides deubiquitination. See, e.g., Hochstrasser (1996) Cell 84:813-815; Hicke and Riezman (1996) Cell 84:277-287; and Chen, et al. (1996) Cell 84:853-862.

The MD gene products will be screened for cell signaling activities. See, e.g., Miyake, et al. (1998) J. Immunol.
25 161:1348-1353; Kobe and Deisenhofer (1994) Trends Biochem. Sci. 19:412.

XXI. Antagonizing cyclin E2 proteins

The inhibition of cell cycle progression is especially
30 important for the control of abnormally proliferative diseases, e.g., cancer. Several methods are available to accomplish this control. The ability of cyclin binding is inhibited by the use, e.g., of antibodies raised against the cyclin binding proteins. Other elements include, e.g., peptidomimetics which are peptides
35 designed to mimic the binding site of cyclin associated proteins and disrupt the interaction of these proteins with cyclin. The

most effective method to block cell cycle progression is the use of small molecules, e.g., to block the interaction of the associated proteins with cyclin, or to block downstream activity of the associated proteins, as described, e.g., in Hung, et al. (1996) Chemistry and Biology 3:623-639. Exposure of a cell to these permeable small molecules should cause a conditional loss of function of the target protein.

Also included in this category is the use of gene therapy to block the expression of the cyclin associated protein or gene transcription factors. Methods of using gene therapy are described, e.g., in Goodnow (1992) "Transgenic Animals" in Roitt (ed.) Encyclopedia of Immunology, Academic Press, San Diego, pp. 1502-1504; Travis (1992) Science 256:1392-1394; Kuhn, et al. (1991) Science 254:707-710; Capecchi (1989) Science 244:1288; Robertson (1987) (ed.) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, IRL Press, Oxford; and Rosenberg (1992) J. Clinical Oncology 10:180-199. Also included is the use of antisense RNA in gene therapy to block expression of the target gene, or proper splicing of gene transcripts.

All citations herein are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

WHAT IS CLAIMED IS:

1. An isolated or recombinant antigenic polypeptide comprising:

- 5 a) a plurality of distinct segments, wherein each said segment has identity to at least 12 contiguous amino acids from the mature SEQ ID NO: 2; or
- b) at least 17 contiguous amino acids from the mature SEQ ID NO: 2.

10 2. The polypeptide of Claim 1, wherein said plurality of segments includes

- a) one of at least 19 contiguous amino acids; or
- b) two of at least 15 contiguous amino acids.

15 3. The polypeptide of Claim 1, wherein said polypeptide:

- a) comprises the mature SEQ ID NO: 2;
- b) binds with specificity to a polyclonal antibody which specifically binds to SEQ ID NO: 2; or
- 20 c) said polypeptide:
- i) is a natural allelic variant of SEQ ID NO: 2;
- ii) is at least 30 amino acids in length;
- iii) exhibits at least two non-overlapping epitopes specific for SEQ ID NO: 2;
- 25 iv) is a synthetic polypeptide;
- v) is attached to a solid substrate; or
- vi) is a 5-fold or less conservative substitution from SEQ ID NO: 2.

30 4. A fusion protein comprising first and second portions, said first portion comprising a polypeptide of Claim 1 and said second portion comprising a detectable marker.

35 5. A pharmaceutical composition comprising a sterile polypeptide of Claim 1 in a pharmaceutically acceptable carrier.

 6. An isolated or recombinant polynucleotide encoding a polypeptide of Claim 1.

7. The polynucleotide of Claim 6, which:
- a) comprises the mature polypeptide coding portion of SEQ ID NO: 1; or
- 5 b) encodes the mature SEQ ID NO: 2.
8. The polynucleotide of Claim 6, wherein said polynucleotide is:
- a) a PCR product;
- 10 b) a hybridization probe;
- c) a mutagenesis primer; or
- d) made by chemical synthesis.
9. The polynucleotide of Claim 6, which is:
- 15 a) detectably labeled;
- b) a deoxyribonucleic acid; or
- c) double stranded.
10. An expression vector comprising a polynucleotide of
- 20 Claim 6.
11. The vector of Claim 10, wherein said polypeptide specifically binds polyclonal antibodies generated against an immunogen of mature SEQ ID NO: 2.
- 25 12. The vector of Claim 10, which
- a) selectively hybridizes under stringent hybridization conditions to a target polynucleotide sequence having at least 60 contiguous nucleotides from SEQ ID NO: 1;
- 30 b) encodes a polypeptide having at least 50 contiguous amino acid residues from mature SEQ ID NO: 2; or
- c) is suitable for transfection into a prokaryote or eukaryote host cell.
- 35 13. The vector of Claim 12, wherein said host cell is:
- a) a mammalian cell;
- b) a bacterial cell;

- c) an insect cell;
- d) a prokaryote;
- e) a eukaryote; or
- f) a COS cell.

5

14. A method of making a polypeptide comprising expressing said vector of Claim 13 in said host cell.

10 15. An isolated or recombinant polynucleotide which hybridizes to the coding portion of SEQ ID NO: 1 under stringent hybridization and wash conditions of at least 50° C, a salt concentration of less than 400 mM, and 50% formamide.

15 16. An expression vector comprising the polynucleotide of Claim 15.

17. The vector of Claim 16 which hybridizes to the coding portion of SEQ ID NO: 1 under stringent hybridization and wash conditions of at least 60° C, a salt concentration of less than 200 mM, and 50% formamide.

18. The vector of Claim 25, which encodes a polypeptide which specifically binds an antibody generated against a mature SEQ ID NO: 2.

25

19. The polynucleotide of Claim 15 which hybridizes to SEQ ID NO: 1, wherein said polynucleotide is:

- a) a PCR product;
- b) a hybridization probe;
- 30 c) a mutagenesis primer; or
- d) made by chemical synthesis.

20. A method of modulating the physiology or development of a cell, comprising contacting said cell with an agonist or 35 antagonist of a polypeptide of Claim 1.

21. A method of detecting the presence of a complementary polynucleotide in a sample, comprising contacting a polynucleotide of Claim 6 that selectively hybridizes with said complementary polynucleotide in said sample to form a detectable duplex; thereby
5 indicating the presence of said polynucleotide in said sample.

22. A method for identifying a compound that binds to a polypeptide of Claim 1, comprising:

- 10 a) incubating components comprising said compound and said polypeptide under conditions sufficient to allow the components to interact; and
- b) measuring the binding of the compound to said polypeptide.

15 23. An isolated or recombinant polynucleotide encoding an antigenic polypeptide comprising:

- a) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 6;
- 20 b) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 8;
- c) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 10;
- d) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 12;
- 25 e) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 17;
- f) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 19;
- 30 g) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 21; or
- h) at least 17 contiguous amino acids from the mature polypeptide from SEQ ID NO: 23.

24. The polynucleotide of Claim 23, encoding all of the polypeptide of:

- 35 a) signal processed SEQ ID NO: 6;
- b) signal processed SEQ ID NO: 8;
- c) signal processed SEQ ID NO: 10;

- 5 d) signal processed SEQ ID NO: 12;
 e) signal processed SEQ ID NO: 17;
 f) SEQ ID NO: 19;
 g) SEQ ID NO: 21; or
 h) SEQ ID NO: 23.

25. The polynucleotide of Claim 23, which hybridizes at 55° C, less than 500 mM salt, and 50% formamide to the:

- 10 a) mature protein coding portion of SEQ ID NO: 5;
 b) signal processed coding portion of SEQ ID NO: 7;
 c) signal processed coding portion of SEQ ID NO: 9;
 d) signal processed coding portion of SEQ ID NO: 11;
 e) mature protein coding portion of SEQ ID NO: 16;
15 f) polypeptide coding portion of SEQ ID NO: 18;
 g) polypeptide coding portion of SEQ ID NO: 20; or
 h) polypeptide coding portion of SEQ ID NO: 22.

26. The polynucleotide of Claim 25, comprising at least 35 contiguous nucleotides of:

- 20 a) mature protein coding portion of SEQ ID NO: 5;
 b) signal processed coding portion of SEQ ID NO: 7;
 c) signal processed coding portion of SEQ ID NO: 9;
 d) signal processed coding portion of SEQ ID NO: 11;
 e) mature protein coding portion of SEQ ID NO: 16;
25 f) polypeptide coding portion of SEQ ID NO: 18;
 g) polypeptide coding portion of SEQ ID NO: 20; or
 h) polypeptide coding portion of SEQ ID NO: 22.

30 27. An expression vector comprising the polynucleotide of Claim 23.

28. A host cell containing the expression vector of Claim 27, including a eukaryotic cell.

35 29. A method of making an antigenic polypeptide comprising expressing a recombinant polynucleotide of Claim 23.

30. A method for detecting a polynucleotide of Claim 23, comprising contacting said polynucleotide with a probe that hybridizes, under stringent conditions, to at least 25 contiguous nucleotides of the:

- 5 a) mature protein coding portion of SEQ ID NO: 5;
 - b) signal processed coding portion of SEQ ID NO: 7;
 - c) signal processed coding portion of SEQ ID NO: 9;
 - d) signal processed coding portion of SEQ ID NO: 11;
 - e) mature protein coding portion of SEQ ID NO: 16;
 - 10 f) polypeptide coding portion of SEQ ID NO: 18;
 - g) polypeptide coding portion of SEQ ID NO: 20; or
 - h) polypeptide coding portion of SEQ ID NO: 22;
- to form a duplex, wherein detection of said duplex indicates the presence of said polynucleotide.

31. A kit for the detection of a polynucleotide of Claim 23, comprising a compartment containing a probe that hybridizes, under stringent hybridization conditions, to at least 17 contiguous nucleotides of a polynucleotide of Claim b1 to form a duplex.

32. The kit of Claim 31, wherein said probe is detectably labeled.

33. A binding compound comprising an antibody binding site which specifically binds to a polypeptide comprising at least 17 contiguous amino acids from:

- a) signal processed SEQ ID NO: 6;
- b) signal processed SEQ ID NO: 8;
- c) signal processed SEQ ID NO: 10;
- 30 d) signal processed SEQ ID NO: 12;
- e) signal processed SEQ ID NO: 17;
- f) SEQ ID NO: 19;
- g) SEQ ID NO: 21; or
- h) SEQ ID NO: 23.

34. The binding compound of Claim 33, wherein:

- a) said antibody binding site is:

- 5
- 1) selectively immunoreactive with the:
- a) signal processed SEQ ID NO: 6;
 - b) signal processed SEQ ID NO: 8;
 - c) signal processed SEQ ID NO: 10;
 - d) signal processed SEQ ID NO: 12;
 - e) signal processed SEQ ID NO: 17;
 - f) SEQ ID NO: 19;
 - g) SEQ ID NO: 21; or
 - h) SEQ ID NO: 23;
- 10
- 2) raised against a purified or recombinantly produced human HDTEA84 protein;
- 3) raised against a purified or recombinantly produced human HSLJD37R protein; or
- 4) in a monoclonal antibody, Fab, or F(ab)₂; or
- 15
- b) said binding compound is:
- 1) an antibody molecule;
 - 2) a polyclonal antiserum;
 - 3) detectably labeled;
 - 4) sterile; or
 - 20
 - 5) in a buffered composition.

35. A method using the binding compound of Claim 33, comprising contacting said binding compound with a biological sample comprising an antigen, thereby forming a binding compound:antigen complex.

25

36. The method of Claim 35, wherein said biological sample is from a human, and wherein said binding compound is an antibody.

37. A detection kit comprising said binding compound of Claim 34, and:

30

- a) instructional material for the use of said binding compound for said detection; or
 - b) a compartment providing segregation of said binding compound.
- 35

38. A substantially pure or isolated antigenic polypeptide, which binds to said binding composition of Claim 33, and further comprises at least 17 contiguous amino acids from:

- a) signal processed SEQ ID NO: 6;
- 5 b) signal processed SEQ ID NO: 8;
- c) signal processed SEQ ID NO: 10;
- d) signal processed SEQ ID NO: 12;
- e) signal processed SEQ ID NO: 17;
- f) SEQ ID NO: 19;
- 10 g) SEQ ID NO: 21; or
- h) SEQ ID NO: 23.

39. The polypeptide of Claim 38, which:

- a) comprises at least a fragment of at least 25 contiguous amino acid residues from a primate HDTEA84 protein;
- 15 b) comprises at least a fragment of at least 25 contiguous amino acid residues from a primate HSLJD37R protein;
- c) comprises at least a fragment of at least 25 contiguous amino acid residues from a rodent or primate RANKL protein;
- 20 d) is a soluble polypeptide;
- e) is detectably labeled;
- f) is in a sterile composition;
- g) is in a buffered composition;
- 25 h) binds to an sialic acid residue;
- i) is recombinantly produced, or
- j) has a naturally occurring polypeptide sequence.

40. The polypeptide of Claim 39, which comprises at least 17 contiguous amino acids from the:

- 30 a) signal processed SEQ ID NO: 6;
- b) signal processed SEQ ID NO: 8;
- c) signal processed SEQ ID NO: 10;
- d) signal processed SEQ ID NO: 12;
- e) signal processed SEQ ID NO: 17;
- 35 f) SEQ ID NO: 19;
- g) SEQ ID NO: 21; or
- h) SEQ ID NO: 23.

41. A method of modulating a precursor cell physiology or function comprising a step of contacting said cell with:

- a) a binding compound which binds to said polypeptide of Claim 38;
- b) an HDTEA84 polypeptide;
- c) an HSLJD37R polypeptide; or
- d) a RANKL polypeptide.

42. The method of Claim 41, wherein said contacting is in combination with a TNF family ligand, or an antagonist of said TNF family ligand.

43. A composition of matter selected from:

- a) a substantially pure or recombinant HCC5 polypeptide exhibiting identity over a length of at least 12 amino acids to SEQ ID NO: 25;
- b) an isolated natural sequence HCC5 of mature SEQ ID NO: 25;
- c) a fusion protein comprising HCC5 sequence;
- d) a substantially pure or recombinant Dub11 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 32 or 34;
- e) an isolated natural sequence Dub11 of mature SEQ ID NO: 32 or 34;
- f) a fusion protein comprising Dub11 sequence;
- g) a substantially pure or recombinant Dub12 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 36 or 38;
- h) an isolated natural sequence Dub12 of mature SEQ ID NO: 36 or 38;
- i) a fusion protein comprising Dub12 sequence;
- j) a substantially pure or recombinant MD-1 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 42;
- k) an isolated natural sequence MD-1 of mature SEQ ID NO: 42;

- 1) a fusion protein comprising primate MD-1 sequence;
m) a substantially pure or recombinant MD-2 polypeptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 44 or 46;
5 n) an isolated natural sequence MD-2 of mature SEQ ID NO: 44 or 46;
o) a fusion protein comprising primate MD-2 sequence;
p) a substantially pure or recombinant MD-2 polypeptide exhibiting identity over a length of at least about 12
10 amino acids to SEQ ID NO: 48 or 49;
q) an isolated natural sequence MD-2 of mature SEQ ID NO: 48; or
r) a fusion protein comprising murine MD-2 sequence.
- 15 44. The composition of Claim 43, which is a substantially pure or isolated:
a) a HCC5 polypeptide, wherein said length is at least 17 amino acids;
b) a Dub11 polypeptide, wherein said length is at least 17
20 amino acids;
c) a Dub12 polypeptide, wherein said length is at least 17 amino acids;
d) a primate MD-1 polypeptide, wherein said length is at least 17 amino acids;
25 e) a primate MD-2 polypeptide, wherein said length is at least 17 amino acids; or
f) a rodent MD-2 polypeptide, wherein said length is at least 17 amino acids.
- 30 45. The composition of Claim 44, which is a substantially pure or isolated:
a) a HCC5 polypeptide, wherein said length is at least 21 amino acids;
b) a Dub11 polypeptide, wherein said length is at least 21
35 amino acids;
c) a Dub12 polypeptide, wherein said length is at least 21 amino acids;

- d) a primate MD-1 polypeptide, wherein said length is at least 21 amino acids;
- e) a primate MD-2 polypeptide, wherein said length is at least 21 amino acids; and
- 5 f) a rodent MD-2 polypeptide, wherein said length is at least 21 amino acids.
46. The composition of matter of Claim 43, wherein said:
- a) HCC5 polypeptide:
- 10 i) is from a primate, including a human;
- ii) comprises at least one polypeptide segment of SEQ ID NO: 25;
- iii) exhibits a plurality of portions exhibiting said identity;
- 15 iv) is a natural allelic variant of HCC5;
- v) has a length at least about 30 amino acids;
- vi) exhibits at least two non-overlapping epitopes which are specific for a primate HCC5;
- vii) exhibits a sequence identity over a length of at least 35 amino acids to a HCC5;
- 20 viii) is glycosylated;
- ix) is a synthetic polypeptide;
- x) is attached to a solid substrate;
- xi) is conjugated to another chemical moiety;
- 25 xii) is a 5-fold or less substitution from natural sequence; or
- xiii) is a deletion or insertion variant from a natural sequence;
- b) Dub11 polypeptide:
- 30 i) is from a primate, including a human;
- ii) comprises at least one polypeptide segment of SEQ ID NO: 32 or 34;
- iii) exhibits a plurality of portions exhibiting said identity;
- 35 iv) is a natural allelic variant of Dub11;
- v) has a length at least about 30 amino acids;

- vi) exhibits at least two non-overlapping epitopes which are specific for a primate Dub11;
- vii) exhibits a sequence identity over a length of at least about 35 amino acids to a Dub11;
- viii) is glycosylated;
- ix) is a synthetic polypeptide;
- x) is attached to a solid substrate;
- xi) is conjugated to another chemical moiety;
- xii) is a 5-fold or less substitution from natural sequence; or
- xiii) is a deletion or insertion variant from a natural sequence;

c) Dub12 polypeptide:

- i) is from a primate, including a human;
- ii) comprises at least one polypeptide segment of SEQ ID NO: 36 or 38;
- iii) exhibits a plurality of portions exhibiting said identity;
- iv) is a natural allelic variant of Dub12;
- v) has a length at least about 30 amino acids;
- vi) exhibits at least two non-overlapping epitopes which are specific for a primate Dub12;
- vii) exhibits a sequence identity over a length of at least about 35 amino acids to a Dub12;
- viii) is glycosylated;
- ix) is a synthetic polypeptide;
- x) is attached to a solid substrate;
- xi) is conjugated to another chemical moiety;
- xii) is a 5-fold or less substitution from natural sequence; or
- xiii) is a deletion or insertion variant from a natural sequence;

d) primate MD-1 polypeptide:

- i) is from a human;
- ii) comprises at least one polypeptide segment of SEQ ID NO: 42;

- iii) exhibits a plurality of portions exhibiting said identity;
- iv) is a natural allelic variant of primate MD-1;
- v) has a length at least about 30 amino acids;
- 5 vi) exhibits at least two non-overlapping epitopes which are specific for a primate MD-1;
- vii) exhibits a sequence identity over a length of at least about 35 amino acids to a primate MD-1;
- viii) is glycosylated;
- 10 ix) is a synthetic polypeptide;
- x) is attached to a solid substrate;
- xi) is conjugated to another chemical moiety;
- xii) is a 5-fold or less substitution from natural sequence; or
- 15 xiii) is a deletion or insertion variant from a natural sequence;
- e) primate MD-2 polypeptide:::
 - i) is from a human;
 - ii) comprises at least one polypeptide segment of SEQ ID NO: 44 or 46;
 - 20 iii) exhibits a plurality of portions exhibiting said identity;
 - iv) is a natural allelic variant of primate MD-2;
 - v) has a length at least about 30 amino acids;
 - 25 vi) exhibits at least two non-overlapping epitopes which are specific for a primate MD-2;
 - vii) exhibits a sequence identity over a length of at least about 35 amino acids to a primate MD-2;
 - viii) is glycosylated;
 - 30 ix) is a synthetic polypeptide;
 - x) is attached to a solid substrate;
 - xi) is conjugated to another chemical moiety;
 - xii) is a 5-fold or less substitution from natural sequence; or
 - 35 xiii) is a deletion or insertion variant from a natural sequence; or
- f) rodent MD-2 polypeptide:

- i) is from a mouse;
- ii) comprises at least one polypeptide segment of SEQ ID NO: 48 or 49;
- iii) exhibits a plurality of portions exhibiting said identity;
- iv) is a natural allelic variant of rodent MD-2;
- v) has a length at least about 30 amino acids;
- vi) exhibits at least two non-overlapping epitopes which are specific for a rodent MD-2;
- vii) exhibits a sequence identity over a length of at least about 35 amino acids to a rodent MD-2;
- viii) is glycosylated;
- ix) is a synthetic polypeptide;
- x) is attached to a solid substrate;
- xi) is conjugated to another chemical moiety;
- xii) is a 5-fold or less substitution from natural sequence; or
- xiii) is a deletion or insertion variant from a natural sequence.

47. A composition comprising a sterile polypeptide of Claim 43, wherein said polypeptide is:

- a) HCC5 polypeptide;
- b) Dub11 polypeptide;
- c) Dub12 polypeptide;
- d) MD-1 polypeptide; or
- e) MD-2 polypeptide.

48. A composition of Claim 43 comprising:

- a) said HCC5 polypeptide and:
 - 1) a carrier, wherein said carrier is:
 - a) an aqueous compound, including water, saline, and/or buffer; and/or
 - b) formulated for oral, rectal, nasal, topical, or parenteral administration;
 - 2) another chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4; or

- 3) an antibody antagonist for a chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4;
- b) said Dub11 polypeptide and a carrier, wherein said carrier is
- 5 a) an aqueous compound, including water, saline, and/or buffer; and/or
- b) formulated for oral, rectal, nasal, topical, or parenteral administration;
- 10 c) said Dub12 polypeptide and a carrier, wherein said carrier is:
- a) an aqueous compound, including water, saline, and/or buffer; and/or
- b) formulated for oral, rectal, nasal, topical, or parenteral administration;
- 15 d) said MD-1 polypeptide and a carrier, wherein said carrier is:
- a) an aqueous compound, including water, saline, and/or buffer; and/or
- 20 b) formulated for oral, rectal, nasal, topical, or parenteral administration;
- e) said MD-2 polypeptide and a carrier, wherein said carrier is:
- a) an aqueous compound, including water, saline, and/or buffer; and/or
- 25 b) formulated for oral, rectal, nasal, topical, or parenteral administration.
49. The fusion protein of Claim 43 comprising:
- 30 a) mature protein sequence of Table 7;
- b) mature protein sequence of Table 9;
- b) mature protein sequence of Table 11;
- c) a detection or purification tag, including a FLAG, His6, or Ig sequence; or
- 35 d) sequence of another chemokine protein with said protein in a).

50. A kit comprising a polypeptide of Claim 43, and:
- a) a compartment comprising said polypeptide; and/or
 - b) instructions for use or disposal of reagents in said kit.

5 51. A binding compound comprising an antigen binding portion from an antibody, which specifically binds to a natural:

- a) HCC5 polypeptide of Claim 43, wherein said antibody:
 - i) is raised against a peptide sequence of a mature polypeptide sequence of Table 7;
 - 10 ii) is raised against a mature HCC5;
 - iii) is raised to a purified HCC5;
 - iv) is immunoselected;
 - v) is a polyclonal antibody;
 - vi) binds to a denatured HCC5; or
 - 15 vii) exhibits a Kd to antigen of at least 30 μ M;
- b) Dub11 polypeptide of Claim 43, wherein said antibody:
 - i) is raised against a peptide sequence of a mature polypeptide sequence of Table 9;
 - 20 ii) is raised against a mature Dub11;
 - iii) is raised to a purified Dub11;
 - iv) is immunoselected;
 - v) is a polyclonal antibody;
 - vi) binds to a denatured Dub11; or
 - vii) exhibits a Kd to antigen of at least 30 μ M;
- 25 c) Dub12 polypeptide of Claim 43, wherein said antibody:
 - i) is raised against a peptide sequence of a mature polypeptide sequence of Table 9;
 - ii) is raised against a mature Dub12;
 - iii) is raised to a purified Dub12;
 - 30 iv) is immunoselected;
 - v) is a polyclonal antibody;
 - vi) binds to a denatured Dub12; or
 - vii) exhibits a Kd to antigen of at least 30 μ M;
- 35 d) a primate MD-1 polypeptide of Claim 43, wherein said antibody:

- 5 i) is raised against a peptide sequence of a mature polypeptide sequence of Table 11;
 ii) is raised against a mature MD-1;
 iii) is raised to a purified MD-1;
 iv) is immunoselected;
 v) is a polyclonal antibody;
 vi) binds to a denatured MD-1; or
 vii) exhibits a Kd to antigen of at least 30 μ M;
- 10 e) a primate MD-2 polypeptide of Claim 43, wherein said antibody:
 i) is raised against a peptide sequence of a mature polypeptide sequence of Table 11;
 ii) is raised against a mature MD-2;
 iii) is raised to a purified MD-2;
15 iv) is immunoselected;
 v) is a polyclonal antibody;
 vi) binds to a denatured MD-2; or
 vii) exhibits a Kd to antigen of at least 30 μ M; or
- 20 f) a rodent MD-2 polypeptide of Claim 43, wherein said antibody:
 i) is raised against a peptide sequence of a mature polypeptide sequence of Table 11;
 ii) is raised against a mature rodent MD-2;
 iii) is raised to a purified rodent MD-2;
25 iv) is immunoselected;
 v) is a polyclonal antibody;
 vi) binds to a denatured rodent MD-2; or
 vii) exhibits a Kd to antigen of at least 30 μ M.
- 30 52. The binding composition of Claim 51, wherein:
 a) said polypeptide is from a primate or rodent;
 b) said binding compound is an Fv, Fab, or Fab2 fragment;
 c) said binding compound is conjugated to another chemical moiety;
35 d) is attached to a solid substrate, including a bead or plastic membrane;

- e) is in a sterile composition; or
- f) is detectably labeled, including a radioactive or fluorescent label.

- 5 53. A kit comprising said binding compound of Claim 51, and:
a) a compartment comprising said binding compound;
b) a compartment comprising purified antigen; and/or
c) instructions for use or disposal of reagents in said kit.
- 10 54. A method of producing an antigen:antibody complex,
comprising contacting an antibody of Claim 51 and:
a) a primate HCC5 polypeptide;
b) a primate Dub11 polypeptide;
c) a primate Dub12 polypeptide;
15 d) a primate MD-1 polypeptide;
e) a primate MD-2 polypeptide; or
f) a rodent MD-2 polypeptide;
thereby allowing said complex to form.
- 20 55. A composition comprising said binding compound of Claim
51 and:
1) a carrier, wherein said carrier is:
a) an aqueous compound, including water, saline, and/or
buffer; and/or
25 b) formulated for oral, rectal, nasal, topical, or
parenteral administration; or
2) an antibody antagonist for another chemokine, including
one selected from the group of HCC1, HCC2, HCC3, and
HCC4.
- 30 56. An isolated or recombinant nucleic acid encoding a
polypeptide or fusion protein of Claim 43, wherein:
A) said HCC5 :
a) polypeptide is from a primate, including a human; or
35 b) nucleic acid:
i) encodes an antigenic HCC5 peptide sequence of Table
7;

- ii) encodes a plurality of antigenic peptide sequences of Table 7;
- iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said HCC5 segment;
- 5 iv) is a hybridization probe for a gene encoding said HCC5 polypeptide; or
- v) further encodes another chemokine, including one selected from the group of HCC1, HCC2, HCC3, and HCC4;
- 10 B) said Dub11:
 - a) polypeptide is from a primate, including a human; or
 - b) nucleic acid:
 - i) encodes an antigenic Dub11 peptide sequence of Table 9;
 - 15 ii) encodes a plurality of antigenic peptide sequences of Table 9;
 - iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said Dub11 segment; or
 - iv) is a hybridization probe for a gene encoding said Dub11 polypeptide;
- 20 C) said Dub12:
 - a) polypeptide is from a primate, including a human; or
 - b) nucleic acid:
 - i) encodes an antigenic Dub12 peptide sequence of Table 9;
 - 25 ii) encodes a plurality of antigenic peptide sequences of Table 9;
 - iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said Dub12 segment;
 - 30 iv) is a hybridization probe for a gene encoding said Dub12 polypeptide;
- D) said primate MD-1:
 - a) polypeptide is from a primate, including a human; or
 - b) nucleic acid:
 - 35 i) encodes an antigenic MD-1 peptide sequence of Table 11;

- ii) encodes a plurality of antigenic peptide sequences of Table 11;
- iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said MD-1 segment;
- 5 iv) is a hybridization probe for a gene encoding said Dub11 polypeptide;
- E) said primate MD-2:
 - a) polypeptide is from a human; or
 - b) nucleic acid:
 - 10 i) encodes an antigenic MD-2 peptide sequence of Table 11;
 - ii) encodes a plurality of antigenic peptide sequences of Table 11;
 - 15 iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said MD-2 segment;
 - iv) is a hybridization probe for a gene encoding said primate MD-2 polypeptide; or
- F) said rodent MD-2:
 - a) polypeptide is from a mouse; or
 - 20 b) nucleic acid:
 - i) encodes an antigenic MD-2 peptide sequence of Table 11;
 - ii) encodes a plurality of antigenic peptide sequences of Table 11;
 - 25 iii) exhibits identity over at least 25 nucleotides to a natural cDNA encoding said MD-2 segment; or
 - iv) is a hybridization probe for a gene encoding said rodent MD-2 polypeptide.
- 30 57. The nucleic acid of Claim 56, which:
 - a) is an expression vector;
 - b) further comprises an origin of replication;
 - c) is from a natural source;
 - d) comprises a detectable label;
 - 35 e) comprises synthetic nucleotide sequence;
 - f) is less than 6 kb, preferably less than 3 kb;
 - g) is from a primate, including a human;

- h) comprises a natural full length coding sequence; or
- i) is a PCR primer, PCR product, or mutagenesis primer.

58. A cell or tissue comprising a recombinant nucleic acid
5 of Claim 56, including wherein said cell is:

- a) a prokaryotic cell;
- b) a eukaryotic cell;
- c) a bacterial cell;
- d) a yeast cell;
- 10 e) an insect cell;
- f) a mammalian cell;
- g) a mouse cell;
- h) a primate cell; or
- i) a human cell.

15

59. A kit comprising said nucleic acid of Claim 56, and:

- a) a compartment comprising said nucleic acid;
- b) a compartment comprising a nucleic acid encoding another
chemokine, including HCC1, HCC2, HCC3, and HCC4; or
- 20 c) instructions for use or disposal of reagents in said kit..

60. A nucleic acid which:

- a) hybridizes under wash conditions of 45° C and less than
2M salt to the polypeptide coding portion of SEQ ID NO:
25 24;
- b) hybridizes under wash conditions of 45° C and less than
2M salt to the polypeptide coding portions of SEQ ID NO:
31 or 33;
- c) hybridizes under wash conditions of 45° C and less than
30 2M salt to the coding portions of SEQ ID NO: 35 or 37;
- d) hybridizes under wash conditions of 45° C and less than
2M salt to the coding portion of SEQ ID NO: 41;
- e) hybridizes under wash conditions of 45° C and less than
2M salt to the coding portion of SEQ ID NO: 43 or 45. or
- 35 f) hybridizes under wash conditions of 45° C and less than
2M salt to the coding portion of SEQ ID NO: 47.

61. The nucleic acid of Claim 57, wherein:

- a) said wash conditions are at 55° C and/or 500 mM salt; or
- b) said wash conditions are at 65° C and/or 150 mM salt.

5 62. A method of modulating physiology or development of a cell or tissue culture cells comprising exposing said cell to an agonist or antagonist of HCC5, primate MD-1, primate MD-2, or rodent MD-2.

10 63. A method of detecting specific binding to a compound, comprising:

- a) contacting said compound to a composition selected from the group of:
 - 15 i) an antigen binding site which specifically binds to a HCC5 chemokine;
 - ii) an antigen binding site which specifically binds to Dub11;
 - iii) an antigen binding site which specifically binds to Dub12;
 - 20 iv) an antigen binding site which specifically binds to primate MD-1;
 - v) an antigen binding site which specifically binds to primate MD-2;
 - vi) an antigen binding site which specifically binds to rodent MD-2;
 - 25 vii) an expression vector encoding a HCC5 chemokine or fragment thereof;
 - viii) an expression vector encoding a Dub11 or fragment thereof;
 - 30 ix) an expression vector encoding a Dub12 or fragment thereof;
 - x) an expression vector encoding a primate MD-1 or fragment thereof;
 - xi) an expression vector encoding a primate MD-2 or fragment thereof;
 - 35 xii) an expression vector encoding a rodent MD-2 or fragment thereof;

- xiii) a substantially pure protein which is specifically recognized by said antigen binding site of (i);
- 5 xiv) a substantially pure protein which is specifically recognized by said antigen binding site of (ii);
- xiv) a substantially pure protein which is specifically recognized by said antigen binding site of (iii);
- xiv) a substantially pure protein which is specifically recognized by said antigen binding site of (iv);
- 10 xiv) a substantially pure protein which is specifically recognized by said antigen binding site of (v);
- xiv) a substantially pure protein which is specifically recognized by said antigen binding site of (vi);
- 15 ix) a substantially pure HCC5 chemokine or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of HCC5 chemokine sequence;
- x) a substantially pure Dub11 or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of Dub11 sequence;
- 20 xi) a substantially pure Dub12 or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of Dub11 sequence;
- xi) a substantially pure primate MD-1 or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of primate MD-1 sequence;
- 25 xi) a substantially pure primate MD-2 or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of primate MD-2 sequence;
- 30 xi) a substantially pure rodent MD-2 or peptide thereof of Claim 43, or a fusion protein comprising a 30 amino acid sequence portion of rodent MD-2 sequence; and
- 35 b) detecting binding of said compound to said composition.
64. An isolated or recombinant polynucleotide which:

- 5
- a) encodes at least 17 contiguous amino acid residues of SEQ ID NO: 54;
 - b) encodes at least two distinct segments of at least 10 contiguous amino acid residues of SEQ ID NO 54; or
 - c) comprises one or more segments at least 21 contiguous nucleotides of SEQ ID NO: 53.

65. A method of making:

- 10
- a) a polypeptide comprising expressing an expression vector of Claim 64, thereby producing said polypeptide;
 - b) a duplex nucleic acid comprising contacting a polynucleotide of Claim 64 with a complementary nucleic acid, thereby resulting in production of said duplex nucleic acid;
 - 15 c) a synthetic polynucleotide of Claim 64, comprising chemically polymerizing nucleotides to produce said polynucleotide; or
 - d) a polynucleotide of Claim 64 comprising using a PCR method.
- 20

66. An isolated or recombinant antigenic polypeptide comprising at least:

- 25
- a) one segment comprising at least 17 contiguous amino acids from SEQ ID NO: 54; or
 - b) at least two distinct segments of at least 11 contiguous amino acids from SEQ ID NO: 54.

67. The antigenic polypeptide of Claim 66, comprising at least one segment comprising at least 17 contiguous amino acids from SEQ ID NO: 54.

30

68. The polypeptide of Claim 66, which exhibits at least two non-overlapping epitopes which are selective for primate protein of SEQ ID NO: 54.

- 35
69. The polypeptide of Claim 66, wherein said polypeptide:
- a) is a 5-fold or less substitution from a natural sequence;
 - or

- b) is a deletion or insertion variant from a natural sequence.

70. A kit comprising said polypeptide of Claim 66, and
5 instructions for the use or disposal of said polypeptide or other reagents of said kit.

71. The antigenic polypeptide of Claim 66, comprising at
least two distinct segments of at least 11 contiguous amino acids
10 from SEQ ID NO: 54.

72. The polypeptide of Claim 71:

- a) which comprises at least one sequence from (SEQ ID NO:
54) KESRYVHD (residues 120-127), DKHFEVLH (residues 127-
15 134), HSDLEPQM (residues 134-141), QKDINKNM (residues 177-184), YAPKLQEF (residues 203-210), SEEDILRM (residues 219-226), LRMELIIL (residues 224-231), ELCPVTTI (residues 237-244), and LFLQVDAL (residues 249-256); and/or
20 b) wherein said segments of at least 11 contiguous amino acids comprise one said segment with at least 14 contiguous amino acids from SEQ ID NO: 54.

73. The polypeptide of Claim 71, which exhibits at least two
25 non-overlapping epitopes which are selective for primate protein of SEQ ID NO: 54.

74. The polypeptide of Claim 71, wherein said polypeptide:

- a) comprises a mature sequence of SEQ ID NO: 2;
30 b) binds with selectivity to an antibody generated against an immunogen of SEQ ID NO: 54;
c) comprises a plurality of polypeptide segments of 17 contiguous amino acids of SEQ ID NO: 54; or
d) is a natural allelic variant of SEQ ID NO: 54.

35

75. The polypeptide of Claim 71, wherein said polypeptide:

- a) is in a sterile composition;

- b) has a length at least 30 amino acids;
- c) is not glycosylated;
- d) is denatured;
- e) is a synthetic polypeptide;
- 5 f) is attached to a solid substrate; or
- g) is a fusion protein with a detection or purification tag, including a FLAG, His6, or Ig sequence.

76. The polypeptide of Claim 71, wherein said
- 10 polypeptide:
- a) is a 5-fold or less substitution from a natural sequence; or
 - b) is a deletion or insertion variant from a natural sequence.

- 15 77. A kit comprising said polypeptide of Claim 71, and instructions for the use or disposal of said polypeptide or other reagents of said kit.

- 20 78. A method using said polypeptide of Claim 71:
- a) to label said polypeptide, comprising labeling said polypeptide with a radioactive label;
 - b) to separate said polypeptide from another polypeptide in a mixture, comprising running said mixture on a
 - 25 chromatography matrix, thereby separating said polypeptides;
 - c) to identify a compound that binds selectively to said polypeptide, comprising incubating said compound with said polypeptide under appropriate conditions; thereby causing said component to bind to said polypeptide; or
 - 30 d) to conjugate said polypeptide to a matrix, comprising derivatizing said polypeptide with a reactive reagent, and conjugating said polypeptide to said matrix; or
 - e) inducing an antibody response to said polypeptide,
 - 35 comprising introducing said polypeptide as an antigen to an immune system, thereby inducing said response.

79. A binding compound comprising an antigen binding portion from an antibody which binds with selectivity to a polypeptide of Claim 66.

5 80. A method of evaluating the selectivity of binding of a compound to cyclin E2, comprising contacting said compound to a recombinant cyclin E2 polypeptide and at least one other cyclin; and comparing binding of said compound to said cyclins.

10 81. The polypeptide of Claim 67:

a) which comprises at least one sequence from (SEQ ID NO: 54) KESRYVHD (residues 120-127), DKHFEVLH (residues 127-134), HSDLEPQM (residues 134-141), QKDINKNM (residues 177-184), YAPKLQEF (residues 203-210), SEEDILRM (residues 219-226), LRMELIIL (residues 224-231), ELCPVTII (residues 237-244), and LFLQVDAL (residues 249-256); and/or

b) wherein said segment comprising at least 17 contiguous amino acids exhibits at least 23 contiguous amino acids from SEQ ID NO: 54.

82. The polypeptide of Claim 67, wherein said polypeptide:

a) comprises a mature sequence of SEQ ID NO: 54;

b) binds with selectivity to an antibody generated against an immunogen of SEQ ID NO: 54;

c) comprises a plurality of polypeptide segments comprising at least 17 contiguous amino acids of SEQ ID NO: 54; or

d) is a natural allelic variant of SEQ ID NO: 54.

83. The polypeptide of Claim 67, wherein said polypeptide:

a) is in a sterile composition;

b) has a length at least 30 amino acids;

c) is not glycosylated;

d) is denatured;

e) is a synthetic polypeptide;

f) is attached to a solid substrate; or

g) is a fusion protein with a detection or purification tag, including a FLAG, His6, or Ig sequence.

84. A method using said polypeptide of Claim 67:

- 5 a) to label said polypeptide, comprising labeling said polypeptide with a radioactive label;
- b) to separate said polypeptide from another polypeptide in a mixture, comprising running said mixture on a chromatography matrix, thereby separating said polypeptides;
- 10 c) to identify a compound that binds selectively to said polypeptide, comprising incubating said compound with said polypeptide under appropriate conditions; thereby causing said component to bind to said polypeptide;
- 15 d) to conjugate said polypeptide to a matrix, comprising derivatizing said polypeptide with a reactive reagent, and conjugating said polypeptide to said matrix; or
- e) inducing an antibody response to said polypeptide, comprising introducing said polypeptide as an antigen to an immune system, thereby inducing said response.
- 20

SEQUENCE LISTING

5 <110> Schering Corporation
 <120> Mammalian Genes; Related Reagents and Methods
 <130> SF0819x(omnibusFF)
 10 <140>
 <141>
 <160> 56
 15 <170> PatentIn Ver. 2.0
 <210> 1
 <211> 4064
 20 <212> DNA
 <213> Unknown
 <220>
 <221> CDS
 25 <222> (264)..(2390)
 <220>
 <223> Description of Unknown Organism: primate
 30 <400> 1
 gtgaccaggg agacaaacac ttggagatac ttggggctga gtttgagcaa gactccctaa 60
 cctgtgtctg gacaagtctg atgtcctgtg tggcccaaga agaactgacc cctgtgtctgg 120
 35 agtccccacc gttattgcat ccctgtgtg gtcacctgc tgctgtctcc aggagcccct 180
 gagaagattt gcctcctctc ccctgctaag ctccaggtcc tgagattgaa ttaggggctg 240
 gagctcactg cactccagca gtc atg gga ccc agg ata ggg cca gcg ggt gag 293
 40 Met Gly Pro Arg Ile Gly Pro Ala Gly Glu
 1 5 10
 gta ccc cag gta cca gac aag gaa acc aaa gcc aca atg ggc aca gaa 341
 45 Val Pro Gln Val Pro Asp Lys Glu Thr Lys Ala Thr Met Gly Thr Glu
 15 20 25
 aac aca cct gga ggc aaa gcc agc cca gac cct cag gac gtg cgg cca 389
 Asn Thr Pro Gly Gly Lys Ala Ser Pro Asp Pro Gln Asp Val Arg Pro
 30 35 40
 50 agt gtg ttc cat aac atc aag ctg ttc gtt ctg tgc cac agc ctg ctg 437
 Ser Val Phe His Asn Ile Lys Leu Phe Val Leu Cys His Ser Leu Leu
 45 50 55
 55 cag ctg gcg cag ctc atg atc tcc ggc tac cta aag agc tcc atc tcc 485
 Gln Leu Ala Gln Leu Met Ile Ser Gly Tyr Leu Lys Ser Ser Ile Ser
 60 65 70
 aca gtg gag aag cgc ttc ggc ctc tcc agc cag acg tgc ggg ctg ctg 533
 60 Thr Val Glu Lys Arg Phe Gly Leu Ser Ser Gln Thr Ser Gly Leu Leu
 75 80 85 90

SUBSTITUTE SHEET (rule 26)

5	gcc tcc ttc aac gag gtg ggg aac aca gcc ttg att gtg ttt gtg agc	581
	Ala Ser Phe Asn Glu Val Gly Asn Thr Ala Leu Ile Val Phe Val Ser	
	95 100 105	
10	tat ttt ggc agc cgg gtg cac cga ccc cga atg att ggc tat ggg gct	629
	Tyr Phe Gly Ser Arg Val His Arg Pro Arg Met Ile Gly Tyr Gly Ala	
	110 115 120	
15	atc ctt gtg gcc ctg gcg ggc ctg ctc atg act ctc ccg cac ttc atc	677
	Ile Leu Val Ala Leu Ala Gly Leu Leu Met Thr Leu Pro His Phe Ile	
	125 130 135	
20	tcg gag cca tac cgc tac gac aac acc agc cct gag gat atg cca cag	725
	Ser Glu Pro Tyr Arg Tyr Asp Asn Thr Ser Pro Glu Asp Met Pro Gln	
	140 145 150	
25	gac ttc aag gct tcc ctg tgc ctg ccc aca acc tcg gcc cca gcc tcg	773
	Asp Phe Lys Ala Ser Leu Cys Leu Pro Thr Thr Ser Ala Pro Ala Ser	
	155 160 165 170	
30	gcc ccc tcc aat ggc aac tgc tca agc tac aca gaa acc cag cat ctg	821
	Ala Pro Ser Asn Gly Asn Cys Ser Ser Tyr Thr Glu Thr Gln His Leu	
	175 180 185	
35	agt gtg gtg ggg atc atg ttc gtg gca cag acc ctg ctg ggc gtg ggc	869
	Ser Val Val Gly Ile Met Phe Val Ala Gln Thr Leu Leu Gly Val Gly	
	190 195 200	
40	ggg gtg ccc att cag ccc ttt ggc atc tcc tac atc gat gac ttt gcc	917
	Gly Val Pro Ile Gln Pro Phe Gly Ile Ser Tyr Ile Asp Asp Phe Ala	
	205 210 215	
45	cac aac agc aac tcg ccc ctc tac ctc ggg atc ctg ttt gca gtg acc	965
	His Asn Ser Asn Ser Pro Leu Tyr Leu Gly Ile Leu Phe Ala Val Thr	
	220 225 230	
50	atg atg ggg cca ggc ctg gcc ttt ggg ctg ggc agc ctc atg ctg cgc	1013
	Met Met Gly Pro Gly Leu Ala Phe Gly Leu Gly Ser Leu Met Leu Arg	
	235 240 245 250	
55	ctt tat gtg gac att aac cag atg cca gaa ggt ggt atc agc ctg acc	1061
	Leu Tyr Val Asp Ile Asn Gln Met Pro Glu Gly Gly Ile Ser Leu Thr	
	255 260 265	
60	ata aag gac ccc cga tgg gtg ggt gcc tgg tgg ctg ggt ttc ctc atc	1109
	Ile Lys Asp Pro Arg Trp Val Gly Ala Trp Trp Leu Gly Phe Leu Ile	
	270 275 280	
65	gct gcc ggt gca gtg gcc ctg gct gcc atc ccc tac ttc ttc ttc ccc	1157
	Ala Ala Gly Ala Val Ala Leu Ala Ala Ile Pro Tyr Phe Phe Phe Pro	
	285 290 295	
70	aag gaa atg ccc aag gaa aaa cgt gag ctt cag ttt cgg cga aag gtc	1205
	Lys Glu Met Pro Lys Glu Lys Arg Glu Leu Gln Phe Arg Arg Lys Val	
	300 305 310	
75	tta gca gtc aca gac tca cct gcc agg aag ggc aag gac tct ccc tct	1253
	Leu Ala Val Thr Asp Ser Pro Ala Arg Lys Gly Lys Asp Ser Pro Ser	
	315 320 325 330	

5	aag cag agc cct ggg gag tcc acg aag aag cag gat ggc cta gtc cag	1301
	Lys Gln Ser Pro Gly Glu Ser Thr Lys Lys Gln Asp Gly Leu Val Gln	
	335 340 345	
10	att gca cca aac ctg act gtg atc cag ttc att aaa gtc ttc ccc agg	1349
	Ile Ala Pro Asn Leu Thr Val Ile Gln Phe Ile Lys Val Phe Pro Arg	
	350 355 360	
15	gtg ctg ctg cag acc cta cgc cac ccc atc ttc ctg ctg gtg gtc ctg	1397
	Val Leu Leu Gln Thr Leu Arg His Pro Ile Phe Leu Leu Val Val Leu	
	365 370 375	
20	tcc cag gta tgc ttg tca tcc atg gct gcg ggc atg gcc acc ttc ctg	1445
	Ser Gln Val Cys Leu Ser Ser Met Ala Ala Gly Met Ala Thr Phe Leu	
	380 385 390	
25	ccc aag ttc ctg gag cgc cag ttt tcc atc aca gcc tcc tac gcc aac	1493
	Pro Lys Phe Leu Glu Arg Gln Phe Ser Ile Thr Ala Ser Tyr Ala Asn	
	395 400 405 410	
30	ctg ctc atc ggc tgc ctc tcc ttc cct tcg gtc atc gtg ggc atc gtg	1541
	Leu Leu Ile Gly Cys Leu Ser Phe Pro Ser Val Ile Val Gly Ile Val	
	415 420 425	
35	gtg ggt ggc gtc ctg gtc aag cgg ctc cac ctg ggc cct gtg gga tgc	1589
	Val Gly Gly Val Leu Val Lys Arg Leu His Leu Gly Pro Val Gly Cys	
	430 435 440	
40	ggt gcc ctt tgc ctg ctg ggg atg ctg ctg tgc ctc ttc ttc agc ctg	1637
	Gly Ala Leu Cys Leu Leu Gly Met Leu Leu Cys Leu Phe Phe Ser Leu	
	445 450 455	
45	ccg ctc ttc ttt atc ggc tgc tcc agc cac cag att gcg ggc atc aca	1685
	Pro Leu Phe Phe Ile Gly Cys Ser Ser His Gln Ile Ala Gly Ile Thr	
	460 465 470	
50	cac cag acc agt gcc cac cct ggg ctg gag ctg tct cca agc tgc atg	1733
	His Gln Thr Ser Ala His Pro Gly Leu Glu Leu Ser Pro Ser Cys Met	
	475 480 485 490	
55	gag gcc tgc tcc tgc cca ttg gac ggc ttt aac cct gtc tgc gac ccc	1781
	Glu Ala Cys Ser Cys Pro Leu Asp Gly Phe Asn Pro Val Cys Asp Pro	
	495 500 505	
60	agc act cgt gtg gaa tac atc aca ccc tgc cac gca ggc tgc tca agc	1829
	Ser Thr Arg Val Glu Tyr Ile Thr Pro Cys His Ala Gly Cys Ser Ser	
	510 515 520	
65	tgg gtg gtc cag gat gct ctg gac aac agc cag gtt ttc tac acc aac	1877
	Trp Val Val Gln Asp Ala Leu Asp Asn Ser Gln Val Phe Tyr Thr Asn	
	525 530 535	
70	tgc agc tgc gtg gtg gag ggc aac ccc gtg ctg gca gga tcc tgc gac	1925
	Cys Ser Cys Val Val Glu Gly Asn Pro Val Leu Ala Gly Ser Cys Asp	
	540 545 550	
75	tca acg tgc agc cat ctg gtg gtg ccc ttc ctg ctc ctg gtc agc ctg	1973
	Ser Thr Cys Ser His Leu Val Val Pro Phe Leu Leu Leu Val Ser Leu	
	555 560 565 570	

SUBSTITUTE SHEET (rule 26)

5

tagggaggga gactcaggcc cacacttggg tatttttctaa tttcagacaa acacacactc 3250
 5 agcgcgcaact cactgattcc tacacattgc caagatttca cacatgtgac cagggggccac 3310
 caaagtcacct gtgacctttg tgactaggat cctaatttct ctattttctc ctgggtgcct 3370
 10 gggctctgtgt cacctggggc agtgtggata atgttttagtt ctgtgacact gttttttggg 3430
 ggtggcacct ggttctccga tgcctgggct ggtgtcaggc ccaggactgt agtgctggga 3490
 gcagtaaagc tcagctctgt gtaatgagt atgctatggc ttgctcgtgt cttatgatcc 3550
 15 aatccttttc tacatcagcc cttgttttgt tttatggcta gtcttatctg gcctggttat 3610
 ttccttgccg ggaggagagg gtttgcta atctgtccag cccaacctat taccacccca 3670
 20 cctcgtggg acctactgct cgggaggcag cagacaggga gccaccagca gtggcttcc 3730
 ggccctgtgc tgggggtggg gggaagctgg gggcacatgt ggcccttgcc ttctgagcag 3790
 ctcccagtgc cagggctttg agactttccc acatgataaa agaaaaggga ggtacagaag 3850
 25 ttccaattcc ctttttattt tgctggttg tatctgtaa tgtttaataa atatctgagc 3910
 atgtatctat caacgccaag aatttcaaag tctccttcaa caatatgagg cttttaggat 3970
 30 gtttatattc cttcatccct cttgtttccc aggttttgca gggaaaaaag tctggaatta 4030
 tagatacagc ttattattaa atttgttctt gcat 4064

35 <210> 2
 <211> 709
 <212> PRT
 <213> Unknown

40 <400> 2
 Met Gly Pro Arg Ile Gly Pro Ala Gly Glu Val Pro Gln Val Pro Asp
 1 5 10 15
 Lys Glu Thr Lys Ala Thr Met Gly Thr Glu Asn Thr Pro Gly Gly Lys
 20 25 30
 Ala Ser Pro Asp Pro Gln Asp Val Arg Pro Ser Val Phe His Asn Ile
 35 40 45
 50 Lys Leu Phe Val Leu Cys His Ser Leu Leu Gln Leu Ala Gln Leu Met
 50 55 60
 Ile Ser Gly Tyr Leu Lys Ser Ser Ile Ser Thr Val Glu Lys Arg Phe
 65 70 75 80
 55 Gly Leu Ser Ser Gln Thr Ser Gly Leu Leu Ala Ser Phe Asn Glu Val
 85 90 95
 60 Gly Asn Thr Ala Leu Ile Val Phe Val Ser Tyr Phe Gly Ser Arg Val
 100 105 110
 His Arg Pro Arg Met Ile Gly Tyr Gly Ala Ile Leu Val Ala Leu Ala
 115 120 125

SUBSTITUTE SHEET (rule 26)

5 Gly Leu Leu Met Thr Leu Pro His Phe Ile Ser Glu Pro Tyr Arg Tyr
 130 135 140
 Asp Asn Thr Ser Pro Glu Asp Met Pro Gln Asp Phe Lys Ala Ser Leu
 145 150 155 160
 10 Cys Leu Pro Thr Thr Ser Ala Pro Ala Ser Ala Pro Ser Asn Gly Asn
 165 170 175
 15 Cys Ser Ser Tyr Thr Glu Thr Gln His Leu Ser Val Val Gly Ile Met
 180 185 190
 Phe Val Ala Gln Thr Leu Leu Gly Val Gly Gly Val Pro Ile Gln Pro
 195 200 205
 20 Phe Gly Ile Ser Tyr Ile Asp Asp Phe Ala His Asn Ser Asn Ser Pro
 210 215 220
 Leu Tyr Leu Gly Ile Leu Phe Ala Val Thr Met Met Gly Pro Gly Leu
 225 230 235 240
 25 Ala Phe Gly Leu Gly Ser Leu Met Leu Arg Leu Tyr Val Asp Ile Asn
 245 250 255
 30 Gln Met Pro Glu Gly Gly Ile Ser Leu Thr Ile Lys Asp Pro Arg Trp
 260 265 270
 Val Gly Ala Trp Trp Leu Gly Phe Leu Ile Ala Ala Gly Ala Val Ala
 275 280 285
 35 Leu Ala Ala Ile Pro Tyr Phe Phe Phe Pro Lys Glu Met Pro Lys Glu
 290 295 300
 Lys Arg Glu Leu Gln Phe Arg Arg Lys Val Leu Ala Val Thr Asp Ser
 305 310 315 320
 40 Pro Ala Arg Lys Gly Lys Asp Ser Pro Ser Lys Gln Ser Pro Gly Glu
 325 330 335
 45 Ser Thr Lys Lys Gln Asp Gly Leu Val Gln Ile Ala Pro Asn Leu Thr
 340 345 350
 Val Ile Gln Phe Ile Lys Val Phe Pro Arg Val Leu Leu Gln Thr Leu
 355 360 365
 50 Arg His Pro Ile Phe Leu Leu Val Val Leu Ser Gln Val Cys Leu Ser
 370 375 380
 Ser Met Ala Ala Gly Met Ala Thr Phe Leu Pro Lys Phe Leu Glu Arg
 385 390 395 400
 55 Gln Phe Ser Ile Thr Ala Ser Tyr Ala Asn Leu Leu Ile Gly Cys Leu
 405 410 415
 60 Ser Phe Pro Ser Val Ile Val Gly Ile Val Val Gly Gly Val Leu Val
 420 425 430
 Lys Arg Leu His Leu Gly Pro Val Gly Cys Gly Ala Leu Cys Leu Leu
 435 440 445

SUBSTITUTE SHEET (rule 26)

5 Gly Met Leu Leu Cys Leu Phe Phe Ser Leu Pro Leu Phe Phe Ile Gly
 450 455 460
 Cys Ser Ser His Gln Ile Ala Gly Ile Thr His Gln Thr Ser Ala His
 465 470 475 480
 10 Pro Gly Leu Glu Leu Ser Pro Ser Cys Met Glu Ala Cys Ser Cys Pro
 485 490 495
 Leu Asp Gly Phe Asn Pro Val Cys Asp Pro Ser Thr Arg Val Glu Tyr
 500 505 510
 15 Ile Thr Pro Cys His Ala Gly Cys Ser Ser Trp Val Val Gln Asp Ala
 515 520 525
 20 Leu Asp Asn Ser Gln Val Phe Tyr Thr Asn Cys Ser Cys Val Val Glu
 530 535 540
 Gly Asn Pro Val Leu Ala Gly Ser Cys Asp Ser Thr Cys Ser His Leu
 545 550 555 560
 25 Val Val Pro Phe Leu Leu Leu Val Ser Leu Gly Ser Ala Leu Ala Cys
 565 570 575
 Leu Thr His Thr Pro Ser Phe Met Leu Ile Leu Arg Gly Val Lys Lys
 580 585 590
 30 Glu Asp Lys Thr Leu Ala Val Gly Ile Gln Phe Met Phe Leu Arg Ile
 595 600 605
 35 Leu Ala Trp Met Pro Ser Pro Val Ile His Gly Ser Ala Ile Asp Thr
 610 615 620
 Thr Cys Val His Trp Ala Leu Ser Cys Gly Arg Arg Ala Val Cys Arg
 625 630 635 640
 40 Tyr Tyr Asn Asn Asp Leu Leu Arg Asn Arg Phe Ile Gly Leu Gln Phe
 645 650 655
 Phe Phe Lys Thr Gly Ser Val Ile Cys Phe Ala Leu Val Leu Ala Val
 660 665 670
 45 Leu Arg Gln Gln Asp Lys Glu Ala Arg Thr Lys Glu Ser Arg Ser Ser
 675 680 685
 50 Pro Ala Val Glu Gln Gln Leu Leu Val Ser Gly Pro Gly Lys Lys Pro
 690 695 700
 Glu Asp Ser Arg Val
 705
 55 <210> 3
 <211> 643
 <212> PRT
 <213> Unknown
 60 <220>
 <223> Description of Unknown Organism:primate

SUBSTITUTE SHEET (rule 26)

<400> 3

5 Met Gly Leu Leu Pro Lys Leu Gly Val Ser Gln Gly Ser Asp Thr Ser
1 5 10 15

Thr Ser Arg Ala Gly Arg Cys Ala Arg Ser Val Phe Gly Asn Ile Lys
20 25 30

10 Val Phe Val Leu Cys Gln Gly Leu Leu Gln Leu Cys Gln Leu Leu Tyr
35 40 45

15 Ser Ala Tyr Phe Lys Ser Ser Leu Thr Thr Ile Glu Lys Arg Phe Gly
50 55 60

Leu Ser Ser Ser Ser Ser Gly Leu Ile Ser Ser Leu Asn Glu Ile Ser
65 70 75 80

20 Asn Ala Ile Leu Ile Ile Phe Val Ser Tyr Phe Gly Ser Arg Val His
85 90 95

Arg Pro Arg Leu Ile Gly Ile Gly Gly Leu Phe Leu Ala Ala Gly Ala
100 105 110

25 Phe Ile Leu Thr Leu Pro His Phe Leu Ser Glu Pro Tyr Gln Tyr Thr
115 120 125

30 Leu Ala Ser Thr Gly Asn Asn Ser Arg Leu Gln Ala Glu Leu Cys Gln
130 135 140

Lys His Trp Gln Asp Leu Pro Pro Ser Lys Cys His Ser Thr Thr Gln
145 150 155 160

35 Asn Pro Gln Lys Glu Thr Ser Ser Met Trp Gly Leu Met Val Val Ala
165 170 175

Gln Leu Leu Ala Gly Ile Gly Thr Val Pro Ile Gln Pro Phe Gly Ile
180 185 190

40 Ser Tyr Val Asp Asp Phe Ser Glu Pro Ser Asn Ser Pro Leu Tyr Ile
195 200 205

45 Ser Ile Leu Phe Ala Ile Ser Val Phe Gly Pro Ala Phe Gly Tyr Leu
210 215 220

Leu Gly Ser Ile Met Leu Gln Ile Phe Val Asp Tyr Gly Arg Val Asn
225 230 235 240

50 Thr Ala Ala Val Asn Leu Val Pro Gly Asp Pro Arg Trp Ile Gly Ala
245 250 255

Trp Trp Leu Gly Leu Leu Ile Ser Ser Ala Leu Leu Val Leu Thr Ser
260 265 270

55 Phe Pro Phe Phe Phe Phe Pro Arg Ala Met Pro Ile Gly Ala Lys Arg
275 280 285

60 Ala Pro Ala Thr Ala Asp Glu Ala Arg Lys Leu Glu Glu Ala Lys Ser
290 295 300

Arg Gly Ser Leu Val Asp Phe Ile Lys Arg Phe Pro Cys Ile Phe Leu
305 310 315 320

SUBSTITUTE SHEET (rule 26)

5 Arg Leu Leu Met Asn Ser Leu Phe Val Leu Val Val Leu Ala Gln Cys
 325 330 335
 Thr Phe Ser Ser Val Ile Ala Gly Leu Ser Thr Phe Leu Asn Lys Phe
 340 345 350
 10 Leu Glu Lys Gln Tyr Gly Thr Ser Ala Ala Tyr Ala Asn Phe Leu Ile
 355 360 365
 Gly Ala Val Asn Leu Pro Ala Ala Ala Leu Gly Met Leu Phe Gly Gly
 370 375 380
 15 Ile Leu Met Lys Arg Phe Val Phe Ser Leu Gln Thr Ile Pro Arg Ile
 385 390 395 400
 20 Ala Thr Thr Ile Ile Thr Ile Ser Met Ile Leu Cys Val Pro Leu Phe
 405 410 415
 Phe Met Gly Cys Ser Thr Pro Thr Val Ala Glu Val Tyr Pro Pro Ser
 420 425 430
 25 Thr Ser Ser Ser Ile His Pro Gln Ser Pro Ala Cys Arg Arg Asp Cys
 435 440 445
 Ser Cys Pro Asp Ser Ile Phe His Pro Val Cys Gly Asp Asn Gly Ile
 450 455 460
 30 Glu Tyr Leu Ser Pro Cys His Ala Gly Cys Ser Asn Ile Asn Met Ser
 465 470 475 480
 35 Ser Ala Thr Ser Lys Gln Leu Ile Tyr Leu Asn Cys Ser Cys Val Thr
 485 490 495
 Gly Gly Ser Ala Ser Ala Lys Thr Gly Ser Cys Pro Val Pro Cys Ala
 500 505 510
 40 His Phe Leu Leu Pro Ala Ile Phe Leu Ile Ser Phe Val Ser Leu Ile
 515 520 525
 Ala Cys Ile Ser His Asn Pro Leu Tyr Met Met Val Leu Arg Val Val
 530 535 540
 45 Asn Gln Glu Glu Lys Ser Phe Ala Ile Gly Val Gln Phe Leu Leu Met
 545 550 555 560
 50 Arg Leu Leu Ala Trp Leu Pro Ser Pro Ala Leu Tyr Gly Leu Thr Ile
 565 570 575
 Asp His Ser Cys Ile Arg Trp Asn Ser Leu Cys Leu Gly Arg Arg Gly
 580 585 590
 55 Ala Cys Ala Tyr Tyr Asp Asn Asp Ala Leu Arg Asp Arg Tyr Leu Gly
 595 600 605
 Leu Gln Met Gly Tyr Lys Ala Leu Gly Met Leu Leu Leu Cys Phe Ile
 610 615 620
 60 Ser Trp Arg Val Lys Lys Asn Lys Glu Tyr Asn Val Gln Lys Ala Ala
 625 630 635 640
 Gly Leu Ile

SUBSTITUTE SHEET (rule 26)

10

5 <210> 4
 <211> 643
 <212> PRT
 <213> Unknown

10 <220>
 <223> Description of Unknown Organism: rodent
 <400> 4

15 Met Gly Leu Leu Leu Lys Pro Gly Ala Arg Gln Gly Ser Gly Thr Ser
 1 5 10 15
 Ser Val Pro Asp Arg Arg Cys Pro Arg Ser Val Phe Ser Asn Ile Lys
 20 25 30
 Val Phe Val Leu Cys His Gly Leu Leu Gln Leu Cys Gln Leu Leu Tyr
 35 40 45
 Ser Ala Tyr Phe Lys Ser Ser Leu Thr Thr Ile Glu Lys Arg Phe Gly
 50 55 60
 25 Leu Ser Ser Ser Ser Ser Gly Leu Ile Ser Ser Leu Asn Glu Ile Ser
 65 70 75 80
 Asn Ala Thr Leu Ile Ile Phe Ile Ser Tyr Phe Gly Ser Arg Val Asn
 85 90 95
 Arg Pro Arg Met Ile Gly Ile Gly Gly Leu Leu Leu Ala Ala Gly Ala
 100 105 110
 35 Phe Val Leu Thr Leu Pro His Phe Leu Ser Glu Pro Tyr Gln Tyr Thr
 115 120 125
 Ser Thr Thr Asp Gly Asn Arg Ser Ser Phe Gln Thr Asp Leu Cys Gln
 130 135 140
 40 Lys His Phe Gly Ala Leu Pro Pro Ser Lys Cys His Ser Thr Val Pro
 145 150 155 160
 Asp Thr His Lys Glu Thr Ser Ser Leu Trp Gly Leu Met Val Val Ala
 165 170 175
 Gln Leu Leu Ala Gly Ile Gly Thr Val Pro Ile Gln Pro Phe Gly Ile
 180 185 190
 50 Ser Tyr Val Asp Asp Phe Ala Glu Pro Thr Asn Ser Pro Leu Tyr Ile
 195 200 205
 Ser Ile Leu Phe Ala Ile Ala Val Phe Gly Pro Ala Phe Gly Tyr Leu
 210 215 220
 55 Leu Gly Ser Val Met Leu Arg Ile Phe Val Asp Tyr Gly Arg Val Asp
 225 230 235 240
 Thr Ala Thr Val Asn Leu Ser Pro Gly Asp Pro Arg Trp Ile Gly Ala
 245 250 255
 60 Trp Trp Leu Gly Leu Leu Ile Ser Ser Gly Phe Leu Ile Val Thr Ser
 260 265 270

SUBSTITUTE SHEET (rule 26)

12

5 Cys Ala Tyr Tyr Asp Asn Asp Ala Leu Arg Asn Arg Tyr Leu Gly Leu
595 600 605

Gln Met Val Tyr Lys Ala Leu Gly Thr Leu Leu Leu Phe Phe Ile Ser
610 615 620

10 Trp Arg Met Lys Lys Asn Arg Glu Tyr Ser Leu Gln Glu Asn Thr Ser
625 630 635 640

Gly Leu Ile

15

<210> 5
<211> 1137
<212> DNA
20 <213> Unknown

<220>
<221> CDS
<222> (99)..(998)

25

<220>
<221> misc_difference
<222> (367)
<223> may be A; translation would be Asn

30

<220>
<223> Description of Unknown Organism:primate

<400> 5

35 cgcaggcgga ccgggggcaa agggaggtggc atgtcgggtca ggcacagcag ggtcctgtgt 60

ccgcgctgag ccgcgctctc cctgctccag caaggacc atg agg gcg ctg gag ggg 116
Met Arg Ala Leu Glu Gly
1 5

40

cca ggc ctg tgc ctg ctg tgc ctg gtg ttg gcg ctg cct gcc ctg ctg 164
Pro Gly Leu Ser Leu Leu Cys Leu Val Leu Ala Leu Pro Ala Leu Leu
10 15 20

45 ccg gtg ccg gct gta cgc gga gtg gca gaa aca ccc acc tac ccc tgg 212
Pro Val Pro Ala Val Arg Gly Val Ala Glu Thr Pro Thr Tyr Pro Trp
25 30 35

50 cgg gac gca gag aca ggg gag cgg ctg gtg tgc gcc cag tgc ccc cca 260
Arg Asp Ala Glu Thr Gly Glu Arg Leu Val Cys Ala Gln Cys Pro Pro
40 45 50

55 ggc acc ttt gtg cag cgg ccg tgc cgc cga gac agc ccc atg acg tgt 308
Gly Thr Phe Val Gln Arg Pro Cys Arg Arg Asp Ser Pro Met Thr Cys
55 60 65 70

ggc ccg tgt cca ccg cgc cac tac acg cag ttc tgg aac tac ctg gag 356
Gly Pro Cys Pro Pro Arg His Tyr Thr Gln Phe Trp Asn Tyr Leu Glu
75 80 85

60 cgc tgc cgc tac tgc tac gtc ctc tgc ggg gag cgt gag gag gag gca 404
Arg Cys Arg Tyr Cys Tyr Val Leu Cys Gly Glu Arg Glu Glu Glu Ala
90 95 100

SUBSTITUTE SHEET (rule 26)

14.

5 <213> Unknown
 <400> 6
 Met Arg Ala Leu Glu Gly Pro Gly Leu Ser Leu Leu Cys Leu Val Leu
 1 5 10 15
 10 Ala Leu Pro Ala Leu Leu Pro Val Pro Ala Val Arg Gly Val Ala Glu
 20 25 30
 Thr Pro Thr Tyr Pro Trp Arg Asp Ala Glu Thr Gly Glu Arg Leu Val
 35 40 45
 15 Cys Ala Gln Cys Pro Pro Gly Thr Phe Val Gln Arg Pro Cys Arg Arg
 50 55 60
 20 Asp Ser Pro Met Thr Cys Gly Pro Cys Pro Pro Arg His Tyr Thr Gln
 65 70 75 80
 Phe Trp Asn Tyr Leu Glu Arg Cys Arg Tyr Cys Tyr Val Leu Cys Gly
 85 90 95
 25 Glu Arg Glu Glu Glu Ala Arg Ala Cys His Ala Thr His Asn Arg Ala
 100 105 110
 Cys Arg Cys Arg Thr Gly Phe Phe Ala His Ala Gly Phe Cys Leu Glu
 115 120 125
 30 His Ala Ser Cys Pro Pro Gly Ala Gly Val Ile Ala Pro Gly Thr Pro
 130 135 140
 35 Ser Gln Asn Thr Gln Cys Gln Pro Cys Pro Pro Gly Thr Phe Ser Ala
 145 150 155 160
 Ser Ser Ser Ser Ser Glu Gln Cys Gln Pro His Arg Asn Cys Thr Ala
 165 170 175
 40 Leu Gly Leu Ala Leu Asn Val Pro Gly Ser Ser Ser His Asp Thr Leu
 180 185 190
 Cys Thr Ser Cys Thr Gly Phe Pro Leu Ser Thr Arg Val Pro Gly Ala
 195 200 205
 45 Glu Glu Cys Glu Arg Ala Val Ile Asp Phe Val Ala Phe Gln Asp Ile
 210 215 220
 50 Ser Ile Lys Arg Leu Gln Arg Leu Leu Gln Ala Leu Glu Ala Pro Glu
 225 230 235 240
 Gly Trp Gly Pro Thr Pro Arg Ala Gly Arg Ala Ala Leu Gln Leu Lys
 245 250 255
 55 Leu Arg Arg Arg Leu Thr Glu Leu Leu Gly Ala Gln Asp Gly Ala Leu
 260 265 270
 Leu Val Arg Leu Leu Gln Ala Leu Arg Val Ala Arg Met Pro Gly Leu
 275 280 285
 60 Glu Arg Ser Val Arg Glu Arg Phe Leu Pro Val His
 290 295 300

SUBSTITUTE SHEET (rule 26)

15

5 <210> 7
 <211> 1031
 <212> DNA
 <213> Unknown

 10 <220>
 <221> CDS
 <222> (402)..(1031)

 15 <220>
 <223> Description of Unknown Organism:primate

 15 <400> 7
 ccgactcant ccctcgccga ccagtcctggg cagcggagga ggggtggttg cagtggctgg 60
 aagcttcgct atgggaagtc gttcctttgc tctctcgcg ccagtcctcc tccttggttc 120
 20 tcctcagccg ctgtcggagg agagcaccg gagacgcgg ctgcagtcgc ggcggcttct 180
 cccgccttg ggcggcgcgc cgtcgggcag gtgctgagcg cccctagagc ctccttgcc 240
 25 gcctccctcc tctgcccggc cgcagcagtg cacatggggt gttggaggta gatgggctcc 300
 cggcccggga ggcggcggtg gatgcggcgc tgggcagaag cagccgccga ttccagctgc 360
 30 cccgcgcgcc cggggcgccc ctgcgagtc cgggttcagc c atg ggg acc tct ccg 416
 Met Gly Thr Ser Pro
 1 5

 35 agc agc agc acc gcc ctc gcc tcc tgc agc cgc atc gcc cgc cga gcc 464
 Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg Ile Ala Arg Arg Ala
 10 15 20

 aca gcc acg atg atc gcg ggc tcc ctt ctc ctg ctt gga ttc ctt agc 512
 Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu Gly Phe Leu Ser
 25 30 35
 40 acc acc aca gct cag cca gaa cag aag gcc tcg aat ctc att ggc aca 560
 Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser Asn Leu Ile Gly Thr
 40 45 50
 45 tac cgc cat gtt gac cgt gcc acc ggc cag gtg cta acc tgt gac aag 608
 Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val Leu Thr Cys Asp Lys
 55 60 65
 50 tgt cca gca gga acc tat gtc tct gag cat tgt acc aac aca agc tgc 656
 Cys Pro Ala Gly Thr Tyr Val Ser Glu His Cys Thr Asn Thr Ser Cys
 70 75 80 85
 55 gcg tct ggc agc agt tgc cct gtg ggg acc ttt acc agg cat gag aat 704
 Ala Ser Gly Ser Ser Cys Pro Val Gly Thr Phe Thr Arg His Glu Asn
 90 95 100
 60 ggc ata gag aaa tgc cat gac tgt agt cag cca tgc cca tgg cca atg 752
 Gly Ile Glu Lys Cys His Asp Cys Ser Gln Pro Cys Pro Trp Pro Met
 105 110 115
 att gag aaa tta cct tgt gct gcc ttg act gac cga gaa tgc act tgc 800
 Ile Glu Lys Leu Pro Cys Ala Ala Leu Thr Asp Arg Glu Cys Thr Cys
 120 125 130

SUBSTITUTE SHEET (rule 26)

5	cca cct ggc atg ttc cag tct aac gct acc tgt gcc ccc cat acg gtg Pro Pro Gly Met Phe Gln Ser Asn Ala Thr Cys Ala Pro His Thr Val 135 140 145	848
10	tgt cct gtg ggt tgg ggt gtg cgg aag aaa ggg aca gag act gag gat Cys Pro Val Gly Trp Gly Val Arg Lys Lys Gly Thr Glu Thr Glu Asp 150 155 160 165	896
15	gtg cgg tgt aag cag tgt gct cgg ggg tac ttc tca gat gtg cct tct Val Arg Cys Lys Gln Cys Ala Arg Gly Tyr Phe Ser Asp Val Pro Ser 170 175 180	944
20	agt gtg atg aac gca aag cat aca cag act gtc tgg atc aga acc tgg Ser Val Met Asn Ala Lys His Thr Gln Thr Val Trp Ile Arg Thr Trp 185 190 195	992
25	ttg gtg atc aag ccg ggg gga cca agg aga cag aca act Leu Val Ile Lys Pro Gly Gly Pro Arg Arg Gln Thr Thr 200 205 210	1031
30	<210> 8 <211> 210 <212> PRT <213> Unknown	
35	<400> 8 Met Gly Thr Ser Pro Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg 1 5 10 15 Ile Ala Arg Arg Ala Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu 20 25 30 Leu Gly Phe Leu Ser Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser 35 40 45	
40	Asn Leu Ile Gly Thr Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val 50 55 60 Leu Thr Cys Asp Lys Cys Pro Ala Gly Thr Tyr Val Ser Glu His Cys 65 70 75 80 Thr Asn Thr Ser Cys Ala Ser Gly Ser Ser Cys Pro Val Gly Thr Phe 85 90 95	
45	Thr Arg His Glu Asn Gly Ile Glu Lys Cys His Asp Cys Ser Gln Pro 100 105 110 Cys Pro Trp Pro Met Ile Glu Lys Leu Pro Cys Ala Ala Leu Thr Asp 115 120 125	
50	Arg Glu Cys Thr Cys Pro Pro Gly Met Phe Gln Ser Asn Ala Thr Cys 130 135 140	
55	Ala Pro His Thr Val Cys Pro Val Gly Trp Gly Val Arg Lys Lys Gly 145 150 155 160 Thr Glu Thr Glu Asp Val Arg Cys Lys Gln Cys Ala Arg Gly Tyr Phe 165 170 175	

17

5 Ser Asp Val Pro Ser Ser Val Met Asn Ala Lys His Thr Gln Thr Val
 180 185 190
 Trp Ile Arg Thr Trp Leu Val Ile Lys Pro Gly Gly Pro Arg Arg Gln
 195 200 205
 10 Thr Thr
 210
 15 <210> 9
 <211> 2877
 <212> DNA
 <213> Unknown
 20 <220>
 <221> CDS
 <222> (410)..(2374)
 <220>
 25 <223> Description of Unknown Organism:primate
 <400> 9
 ggcacgagcc gactcagtc ctcgccgacc agtctgggca gcggaggagg gtggttggca 60
 gtggctggaa gcttcgctat gggaagtcgt tcctttgctc tctcgcgccc agtcctcttc 120
 30 cctggttctc ctcagccgct gtcggaggag agcaccgga gacgcgggct gcagtcgagg 180
 cggcttctcc ccgcctgggc ggccgcgccc ctgggcaggt gctgagcgcc cctagcgcc 240
 35 cccttgccgc ctccctctc tgcccgccg cagcagtgca catggggtgt tggaggtaga 300
 tgggctcccg gcccgagg cgccggtgga tgcggcgctg ggcagaagca gccgccgatt 360
 40 ccagctgccc cgcgcgccc gggcgccct gcgagtcgcc gggtcagcc atg ggg acc 418
 Met Gly Thr
 1
 tct ccg agc agc agc acc gcc ctc gcc tcc tgc agc cgc atc gcc cgc 466
 45 Ser Pro Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg Ile Ala Arg
 5 10 15
 cga gcc aca gcc acg atg atc gcg ggc tcc ctt ctc ctg ctt gga ttc 514
 Arg Ala Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu Gly Phe
 20 25 30 35
 50 ctt agc acc acc aca gct cag cca gaa cag aag gcc tcg aat ctc att 562
 Leu Ser Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser Asn Leu Ile
 40 45 50
 55 ggc aca tac cgc cat gtt gac cgt gcc acc ggc cag gtg cta acc tgt 610
 Gly Thr Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val Leu Thr Cys
 55 60 65
 60 gac aag tgt cca gca gga acc tat gtc tct gag cat tgt acc aac aca 658
 Asp Lys Cys Pro Ala Gly Thr Tyr Val Ser Glu His Cys Thr Asn Thr
 70 75 80

SUBSTITUTE SHEET (rule 26)

5	agc ctg cgc gtc tgc agc agt tgc cct gtg ggg acc ttt acc agg cat 706 Ser Leu Arg Val Cys Ser Ser Cys Pro Val Gly Thr Phe Thr Arg His 85 90 95
10	gag aat ggc ata gag aaa tgc cat gac tgt agt cag cca tgc cca tgg 754 Glu Asn Gly Ile Glu Lys Cys His Asp Cys Ser Gln Pro Cys Pro Trp 100 105 110 115
15	cca atg att gag aaa tta cct tgt gct gcc ttg act gac cga gaa tgc 802 Pro Met Ile Glu Lys Leu Pro Cys Ala Ala Leu Thr Asp Arg Glu Cys 120 125 130
20	act tgc cca cct ggc atg ttc cag tct aac gct acc tgt gcc ccc cat 850 Thr Cys Pro Pro Gly Met Phe Gln Ser Asn Ala Thr Cys Ala Pro His 135 140 145
25	acg gtg tgt cct gtg ggt tgg ggt gtg cgg aag aaa ggg aca gag act 898 Thr Val Cys Pro Val Gly Trp Gly Val Arg Lys Lys Gly Thr Glu Thr 150 155 160
30	gag gat gtg cgg tgt aag cag tgt gct cgg ggt acc ttc tca gat gtg 946 Glu Asp Val Arg Cys Lys Gln Cys Ala Arg Gly Thr Phe Ser Asp Val 165 170 175
35	cct tct agt gtg atg aaa tgc aaa gca tac aca gac tgt ctg agt cag 994 Pro Ser Ser Val Met Lys Cys Lys Ala Tyr Thr Asp Cys Leu Ser Gln 180 185 190 195
40	aac ctg gtg gtg atc aag ccg ggg acc aag gag aca gac aac gtc tgt 1042 Asn Leu Val Val Ile Lys Pro Gly Thr Lys Glu Thr Asp Asn Val Cys 200 205 210
45	ggc aca ctc ccg tcc ttc tcc agc tcc acc tca cct tcc cct ggc aca 1090 Gly Thr Leu Pro Ser Phe Ser Ser Ser Thr Ser Pro Ser Pro Gly Thr 215 220 225
50	gcc atc ttt cca cgc cct gag cac atg gaa acc cat gaa gtc cct tcc 1138 Ala Ile Phe Pro Arg Pro Glu His Met Glu Thr His Glu Val Pro Ser 230 235 240
55	tcc act tat gtt ccc aaa ggc atg aac tca aca gaa tcc aac tct tct 1186 Ser Thr Tyr Val Pro Lys Gly Met Asn Ser Thr Glu Ser Asn Ser Ser 245 250 255
60	gcc tct gtt aga cca aag gta ctg agt agc atc cag gaa ggg aca gtc 1234 Ala Ser Val Arg Pro Lys Val Leu Ser Ser Ile Gln Glu Gly Thr Val 260 265 270 275
65	cct gac aac aca agc tca gca agg ggg aag gaa gac gtg aac aag acc 1282 Pro Asp Asn Thr Ser Ser Ala Arg Gly Lys Glu Asp Val Asn Lys Thr 280 285 290
70	ctc cca aac ctt cag gta gtc aac cac cag caa ggc ccc cac cac aga 1330 Leu Pro Asn Leu Gln Val Val Asn His Gln Gln Gly Pro His His Arg 295 300 305
75	cac atc ctg aag ctg ctg ccg tcc atg gag gcc act ggg ggc gag aag 1378 His Ile Leu Lys Leu Leu Pro Ser Met Glu Ala Thr Gly Gly Glu Lys 310 315 320

5	tcc agc acg ccc atc aag ggc ccc aag agg gga cat cct aga cag aac	1426
	Ser Ser Thr Pro Ile Lys Gly Pro Lys Arg Gly His Pro Arg Gln Asn	
	325 330 335	
10	cta cac aag cat ttt gac atc aat gag cat ttg ccc tgg atg att gtg	1474
	Leu His Lys His Phe Asp Ile Asn Glu His Leu Pro Trp Met Ile Val	
	340 345 350 355	
15	ctt ttc ctg ctg ctg gtg ctt gtg gtg att gtg gtg tgc agt atc cgg	1522
	Leu Phe Leu Leu Leu Val Leu Val Val Ile Val Val Cys Ser Ile Arg	
	360 365 370	
20	aaa agc tcg agg act ctg aaa aag ggg ccc cgg cag gat ccc agt gcc	1570
	Lys Ser Ser Arg Thr Leu Lys Lys Gly Pro Arg Gln Asp Pro Ser Ala	
	375 380 385	
25	att gtg gaa aag gca ggg ctg aag aaa tcc atg act cca acc cag aac	1618
	Ile Val Glu Lys Ala Gly Leu Lys Lys Ser Met Thr Pro Thr Gln Asn	
	390 395 400	
30	cgg gag aaa tgg atc tac tac tgc aat ggc cat ggt atc gat atc ctg	1666
	Arg Glu Lys Trp Ile Tyr Tyr Cys Asn Gly His Gly Ile Asp Ile Leu	
	405 410 415	
35	aag ctt gta gca gcc caa gtg gga agc cag tgg aaa gat atc tat cag	1714
	Lys Leu Val Ala Ala Gln Val Gly Ser Gln Trp Lys Asp Ile Tyr Gln	
	420 425 430 435	
40	ttt ctt tgc aat gcc agt gag agg gag gtt gct gct ttc tcc aat ggg	1762
	Phe Leu Cys Asn Ala Ser Glu Arg Glu Val Ala Ala Phe Ser Asn Gly	
	440 445 450	
45	tac aca gcc gac cac gag cgg gcc tac gca gct ctg cag cac tgg acc	1810
	Tyr Thr Ala Asp His Glu Arg Ala Tyr Ala Ala Leu Gln His Trp Thr	
	455 460 465	
50	atc cgg ggc ccc gag gcc agc ctc gcc cag cta att agc gcc ctg cgc	1858
	Ile Arg Gly Pro Glu Ala Ser Leu Ala Gln Leu Ile Ser Ala Leu Arg	
	470 475 480	
55	cag cac cgg aga aac gat gtt gtg gag aag att cgt ggg ctg atg gaa	1906
	Gln His Arg Arg Asn Asp Val Val Glu Lys Ile Arg Gly Leu Met Glu	
	485 490 495	
60	gac acc acc cag ctg gaa act gac aaa cta gct ctc ccg atg agc ccc	1954
	Asp Thr Thr Gln Leu Glu Thr Asp Lys Leu Ala Leu Pro Met Ser Pro	
	500 505 510 515	
65	agc ccg ctt agc ccg agc ccc atc ccc agc ccc aac gcg aaa ctt gag	2002
	Ser Pro Leu Ser Pro Ser Pro Ile Pro Ser Pro Asn Ala Lys Leu Glu	
	520 525 530	
70	aat tcc gct ctc ctg acg gtg gag cct tcc cca cag gac aag aac aag	2050
	Asn Ser Ala Leu Thr Val Glu Pro Ser Pro Gln Asp Lys Asn Lys	
	535 540 545	
75	ggc ttc ttc gtg gat gag tcg gag ccc ctt ctc cgc tgt gac tct aca	2098
	Gly Phe Phe Val Asp Glu Ser Glu Pro Leu Leu Arg Cys Asp Ser Thr	
	550 555 560	

5	tcc agc ggc tcc tcc gcg ctg agc agg aac ggt tcc ttt att acc aaa Ser Ser Gly Ser Ser Ala Leu Ser Arg Asn Gly Ser Phe Ile Thr Lys 565 570 575	2146
10	gaa aag aag gac aca gtg ttg cgg cag gta cgc ctg gac ccc tgt gac Glu Lys Lys Asp Thr Val Leu Arg Gln Val Arg Leu Asp Pro Cys Asp 580 585 590 595	2194
15	ttg cag cct atc ttt gat gac atg ctc cac ttt cta aat cct gag gag Leu Gln Pro Ile Phe Asp Asp Met Leu His Phe Leu Asn Pro Glu Glu 600 605 610	2242
20	ctg cgg gtg att gaa gag att ccc cag gct gag gac aaa cta gac cgg Leu Arg Val Ile Glu Glu Ile Pro Gln Ala Glu Asp Lys Leu Asp Arg 615 620 625	2290
25	cta ttc gaa att att gga gtc aag agc cag gaa gcc agc cag acc ctc Leu Phe Glu Ile Ile Gly Val Lys Ser Gln Glu Ala Ser Gln Thr Leu 630 635 640	2338
30	ctg gac tct gtt tat agc cat ctt cct gac ctg ctg tagaacatag Leu Asp Ser Val Tyr Ser His Leu Pro Asp Leu Leu 645 650 655	2384
35	ggatactgca ttctggaaat tactcaattt agtggcaggg tggtttttta atttccttct gtgtctgatt tttgttgttt ggggtgtgtg tgtgtgtttg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg ttttaacagag aatatggcca gtgcttgagt tctttctcct tctctctctc tctttttttt ttaaataact cttctgggaa gttggtttat aagcctttgc caggtgtaac tgttgtgaaa taccacaccac taaagttttt taagttccat attttctcca ttttgccctc ttatgtattt tcaagattat tctgtgcact ttaaatttac tcaacttacc ataatgcag tgtgactttt cccacacact ggattgtgag gctcttaact tcttaaaagt ataatggcat cttgtgaatc ctataagcag tctttatgtc tcttaacatt cacacctact ttttaaaaac aaatattatt act	2444 2504 2564 2624 2684 2744 2804 2864 2877
40	<210> 10 <211> 655 <212> PRT <213> Unknown	
45	<400> 10 Met Gly Thr Ser Pro Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg 1 5 10 15 Ile Ala Arg Arg Ala Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu 20 25 30 Leu Gly Phe Leu Ser Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser 35 40 45 Asn Leu Ile Gly Thr Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val 50 55 60	

5	Leu	Thr	Cys	Asp	Lys	Cys	Pro	Ala	Gly	Thr	Tyr	Val	Ser	Glu	His	Cys	65	70	75	80
	Thr	Asn	Thr	Ser	Leu	Arg	Val	Cys	Ser	Ser	Cys	Pro	Val	Gly	Thr	Phe	85	90	95	
10	Thr	Arg	His	Glu	Asn	Gly	Ile	Glu	Lys	Cys	His	Asp	Cys	Ser	Gln	Pro	100	105	110	
	Cys	Pro	Trp	Pro	Met	Ile	Glu	Lys	Leu	Pro	Cys	Ala	Ala	Leu	Thr	Asp	115	120	125	
15	Arg	Glu	Cys	Thr	Cys	Pro	Pro	Gly	Met	Phe	Gln	Ser	Asn	Ala	Thr	Cys	130	135	140	
	Ala	Pro	His	Thr	Val	Cys	Pro	Val	Gly	Trp	Gly	Val	Arg	Lys	Lys	Gly	145	150	155	160
20	Thr	Glu	Thr	Glu	Asp	Val	Arg	Cys	Lys	Gln	Cys	Ala	Arg	Gly	Thr	Phe	165	170	175	
25	Ser	Asp	Val	Pro	Ser	Ser	Val	Met	Lys	Cys	Lys	Ala	Tyr	Thr	Asp	Cys	180	185	190	
	Leu	Ser	Gln	Asn	Leu	Val	Val	Ile	Lys	Pro	Gly	Thr	Lys	Glu	Thr	Asp	195	200	205	
30	Asn	Val	Cys	Gly	Thr	Leu	Pro	Ser	Phe	Ser	Ser	Ser	Thr	Ser	Pro	Ser	210	215	220	
35	Pro	Gly	Thr	Ala	Ile	Phe	Pro	Arg	Pro	Glu	His	Met	Glu	Thr	His	Glu	225	230	235	240
	Val	Pro	Ser	Ser	Thr	Tyr	Val	Pro	Lys	Gly	Met	Asn	Ser	Thr	Glu	Ser	245	250	255	
40	Asn	Ser	Ser	Ala	Ser	Val	Arg	Pro	Lys	Val	Leu	Ser	Ser	Ile	Gln	Glu	260	265	270	
	Gly	Thr	Val	Pro	Asp	Asn	Thr	Ser	Ser	Ala	Arg	Gly	Lys	Glu	Asp	Val	275	280	285	
45	Asn	Lys	Thr	Leu	Pro	Asn	Leu	Gln	Val	Val	Asn	His	Gln	Gln	Gly	Pro	290	295	300	
50	His	His	Arg	His	Ile	Leu	Lys	Leu	Leu	Pro	Ser	Met	Glu	Ala	Thr	Gly	305	310	315	320
	Gly	Glu	Lys	Ser	Ser	Thr	Pro	Ile	Lys	Gly	Pro	Lys	Arg	Gly	His	Pro	325	330	335	
55	Arg	Gln	Asn	Leu	His	Lys	His	Phe	Asp	Ile	Asn	Glu	His	Leu	Pro	Trp	340	345	350	
	Met	Ile	Val	Leu	Phe	Leu	Leu	Leu	Val	Leu	Val	Val	Ile	Val	Val	Cys	355	360	365	
60	Ser	Ile	Arg	Lys	Ser	Ser	Arg	Thr	Leu	Lys	Lys	Gly	Pro	Arg	Gln	Asp	370	375	380	

	<221> CDS	
5	<222> (1)..(1332)	
	<400> 11	
10	atg ggg acc tct ccg agc agc agc acc gcc ctc gcc tcc tgc agc cgc Met Gly Thr Ser Pro Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg	48
	1 5 10 15	
15	atc gcc cgc cga gcc aca gcc acg atg atc gcg ggc tcc ctt ctc ctg Ile Ala Arg Arg Ala Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu	96
	20 25 30	
20	ctt gga ttc ctt agc acc acc aca gct cag cca gaa cag aag gcc tcg Leu Gly Phe Leu Ser Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser	144
	35 40 45	
25	aat ctc att ggc aca tac cgc cat gtt gac cgt gcc acc ggc cag gtg Asn Leu Ile Gly Thr Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val	192
	50 55 60	
30	cta acc tgt gac aag tgt cca gca gga acc tat gtc tct gag cat tgt Leu Thr Cys Asp Lys Cys Pro Ala Gly Thr Tyr Val Ser Glu His Cys	240
	65 70 75 80	
35	acc aac aca agc ctg cgc gtc tgc agc agt tgc cct gtg ggg acc ttt Thr Asn Thr Ser Leu Arg Val Cys Ser Ser Cys Pro Val Gly Thr Phe	288
	85 90 95	
40	acc agg cat gag aat ggc ata gag aaa tgc cat gac tgt agt cag cca Thr Arg His Glu Asn Gly Ile Glu Lys Cys His Asp Cys Ser Gln Pro	336
	100 105 110	
45	tgc cca tgg cca atg att gag aaa tta cct tgt gct gcc ttg act gac Cys Pro Trp Pro Met Ile Glu Lys Leu Pro Cys Ala Ala Leu Thr Asp	384
	115 120 125	
50	cga gaa tgc act tgc cca cct ggc atg ttc cag tct aac gct acc tgt Arg Glu Cys Thr Cys Pro Pro Gly Met Phe Gln Ser Asn Ala Thr Cys	432
	130 135 140	
55	gcc ccc cat acg gtg tgt cct gtg ggt tgg ggt gtg cgg aag aaa ggg Ala Pro His Thr Val Cys Pro Val Gly Trp Gly Val Arg Lys Lys Gly	480
	145 150 155 160	
60	aca gag act gag gat gtg cgg tgt aag cag tgt gct cgg ggt acc ttc Thr Glu Thr Glu Asp Val Arg Cys Lys Gln Cys Ala Arg Gly Thr Phe	528
	165 170 175	
65	tca gat gtg cct tct agt gtg atg aaa tgc aaa gca tac aca gac tgt Ser Asp Val Pro Ser Ser Val Met Lys Cys Lys Ala Tyr Thr Asp Cys	576
	180 185 190	
70	ctg agt cag aac ctg gtg gtg atc aag ccg ggg acc aag gag aca gac Leu Ser Gln Asn Leu Val Val Ile Lys Pro Gly Thr Lys Glu Thr Asp	624
	195 200 205	
75	aac gtc tgt ggc aca ctc ccg tcc ttc tcc agc tcc acc tca cct tcc Asn Val Cys Gly Thr Leu Pro Ser Phe Ser Ser Ser Thr Ser Pro Ser	672
	210 215 220	

24

5	cct ggc aca gcc atc ttt cca cgc cct gag cac atg gaa acc cat gaa Pro Gly Thr Ala Ile Phe Pro Arg Pro Glu His Met Glu Thr His Glu 225 230 235 240	720
10	gtc cct tcc tcc act tat gtt ccc aaa ggc atg aac tca aca gaa tcc Val Pro Ser Ser Thr Tyr Val Pro Lys Gly Met Asn Ser Thr Glu Ser 245 250 255	768
15	aac tct tct gcc tct gtt aga cca aag gta ctg agt agc atc cag gaa Asn Ser Ser Ala Ser Val Arg Pro Lys Val Leu Ser Ser Ile Gln Glu 260 265 270	816
20	ggg aca gtc cct gac aac aca agc tca gca agg ggg aag gaa gac gtg Gly Thr Val Pro Asp Asn Thr Ser Ser Ala Arg Gly Lys Glu Asp Val 275 280 285	864
25	aac aag acc ctc cca aac ctt cag gta gtc aac cac cag caa ggc ccc Asn Lys Thr Leu Pro Asn Leu Gln Val Val Asn His Gln Gln Gly Pro 290 295 300	912
30	cac cac aga cac atc ctg aag ctg ctg ccg tcc atg gag gcc act ggg His His Arg His Ile Leu Lys Leu Leu Pro Ser Met Glu Ala Thr Gly 305 310 315 320	960
35	ggc gag aag tcc agc acg ccc atc aag ggc ccc aag agg gga cat cct Gly Glu Lys Ser Ser Thr Pro Ile Lys Gly Pro Lys Arg Gly His Pro 325 330 335	1008
40	aga cag aac cta cac aag cat ttt gac atc aat gag cat ttg ccc tgg Arg Gln Asn Leu His Lys His Phe Asp Ile Asn Glu His Leu Pro Trp 340 345 350	1056
45	atg att gtg ctt ttc ctg ctg ctg gtg ctt gtg gtg att gtg gtg tgc Met Ile Val Leu Phe Leu Leu Val Leu Val Val Ile Val Val Cys 355 360 365	1104
50	agt atc cgg aaa agc tcg agg act ctg aaa aag ggg ccc cgg cag gat Ser Ile Arg Lys Ser Ser Arg Thr Leu Lys Lys Gly Pro Arg Gln Asp 370 375 380	1152
55	ccc agt gcc att gtg gaa aag gca ggg ctg aag aaa tcc atg act cca Pro Ser Ala Ile Val Glu Lys Ala Gly Leu Lys Lys Ser Met Thr Pro 385 390 395 400	1200
60	acc cag aac cgg gag aaa tgg atc tac tac tgc aat ggc cat gga ccc Thr Gln Asn Arg Glu Lys Trp Ile Tyr Tyr Cys Asn Gly His Gly Pro 405 410 415	1248
65	cat gat gag gag tgg ggg ttg atg gag aga cat att caa gat att tat His Asp Glu Glu Trp Gly Leu Met Glu Arg His Ile Gln Asp Ile Tyr 420 425 430	1296
70	att caa aga agc aat caa gat tca gaa aga tgg ggt tgataatttt Ile Gln Arg Ser Asn Gln Asp Ser Glu Arg Trp Gly 435 440	1342
75	tacttcaccc tgggaggcag catagtgcag tgaaaggat cgatatactg aagcttgtag	1402
80	cagcccaagt gggaagccag tggaaagata tctatcagtt tctttgcaat gccagtgaga	1462
85	gggaggttgc tg	1474

SUBSTITUTE SHEET (rule 26)

25

5
 <210> 12
 <211> 444
 <212> PRT
 <213> Unknown

10
 <400> 12
 Met Gly Thr Ser Pro Ser Ser Ser Thr Ala Leu Ala Ser Cys Ser Arg
 1 5 10 15

15
 Ile Ala Arg Arg Ala Thr Ala Thr Met Ile Ala Gly Ser Leu Leu Leu
 20 25 30
 Leu Gly Phe Leu Ser Thr Thr Thr Ala Gln Pro Glu Gln Lys Ala Ser
 35 40 45

20
 Asn Leu Ile Gly Thr Tyr Arg His Val Asp Arg Ala Thr Gly Gln Val
 50 55 60

25
 Leu Thr Cys Asp Lys Cys Pro Ala Gly Thr Tyr Val Ser Glu His Cys
 65 70 75 80
 Thr Asn Thr Ser Leu Arg Val Cys Ser Ser Cys Pro Val Gly Thr Phe
 85 90 95

30
 Thr Arg His Glu Asn Gly Ile Glu Lys Cys His Asp Cys Ser Gln Pro
 100 105 110
 Cys Pro Trp Pro Met Ile Glu Lys Leu Pro Cys Ala Ala Leu Thr Asp
 115 120 125

35
 Arg Glu Cys Thr Cys Pro Pro Gly Met Phe Gln Ser Asn Ala Thr Cys
 130 135 140

40
 Ala Pro His Thr Val Cys Pro Val Gly Trp Gly Val Arg Lys Lys Gly
 145 150 155 160
 Thr Glu Thr Glu Asp Val Arg Cys Lys Gln Cys Ala Arg Gly Thr Phe
 165 170 175

45
 Ser Asp Val Pro Ser Ser Val Met Lys Cys Lys Ala Tyr Thr Asp Cys
 180 185 190
 Leu Ser Gln Asn Leu Val Val Ile Lys Pro Gly Thr Lys Glu Thr Asp
 195 200 205

50
 Asn Val Cys Gly Thr Leu Pro Ser Phe Ser Ser Ser Thr Ser Pro Ser
 210 215 220

55
 Pro Gly Thr Ala Ile Phe Pro Arg Pro Glu His Met Glu Thr His Glu
 225 230 235 240
 Val Pro Ser Ser Thr Tyr Val Pro Lys Gly Met Asn Ser Thr Glu Ser
 245 250 255

60
 Asn Ser Ser Ala Ser Val Arg Pro Lys Val Leu Ser Ser Ile Gln Glu
 260 265 270
 Gly Thr Val Pro Asp Asn Thr Ser Ser Ala Arg Gly Lys Glu Asp Val
 275 280 285

SUBSTITUTE SHEET (rule 26)

5 Asn Lys Thr Leu Pro Asn Leu Gln Val Val Asn His Gln Gln Gly Pro
 290 295 300
 His His Arg His Ile Leu Lys Leu Leu Pro Ser Met Glu Ala Thr Gly
 305 310 315 320
 10 Gly Glu Lys Ser Ser Thr Pro Ile Lys Gly Pro Lys Arg Gly His Pro
 325 330 335
 Arg Gln Asn Leu His Lys His Phe Asp Ile Asn Glu His Leu Pro Trp
 15 340 345 350
 Met Ile Val Leu Phe Leu Leu Leu Val Leu Val Val Ile Val Val Cys
 355 360 365
 20 Ser Ile Arg Lys Ser Ser Arg Thr Leu Lys Lys Gly Pro Arg Gln Asp
 370 375 380
 Pro Ser Ala Ile Val Glu Lys Ala Gly Leu Lys Lys Ser Met Thr Pro
 385 390 395 400
 25 Thr Gln Asn Arg Glu Lys Trp Ile Tyr Tyr Cys Asn Gly His Gly Pro
 405 410 415
 His Asp Glu Glu Trp Gly Leu Met Glu Arg His Ile Gln Asp Ile Tyr
 30 420 425 430
 Ile Gln Arg Ser Asn Gln Asp Ser Glu Arg Trp Gly
 435 440
 35 <210> 13
 <211> 227
 <212> PRT
 <213> Unknown
 40 <220>
 <223> Description of Unknown Organism: rodent
 <400> 13
 45 Met Ala Pro Ala Ala Leu Trp Val Ala Leu Val Phe Glu Leu Gln Leu
 1 5 10 15
 Trp Ala Thr Gly His Thr Val Pro Ala Gln Val Val Leu Thr Pro Tyr
 20 25 30
 50 Lys Pro Glu Pro Gly Tyr Glu Cys Gln Ile Ser Gln Glu Tyr Tyr Asp
 35 40 45
 Arg Lys Ala Gln Met Cys Cys Ala Lys Cys Pro Pro Gly Gln Tyr Val
 55 50 55 60
 Lys His Phe Cys Asn Lys Thr Ser Asp Thr Val Cys Ala Asp Cys Glu
 65 70 75 80
 60 Ala Ser Met Tyr Thr Gln Val Trp Asn Gln Phe Arg Thr Cys Leu Ser
 85 90 95
 Cys Ser Ser Ser Cys Thr Thr Asp Gln Val Glu Ile Arg Ala Cys Thr
 100 105 110

SUBSTITUTE SHEET (rule 26)

5 Lys Gln Gln Asn Arg Val Cys Ala Cys Glu Ala Gly Arg Tyr Cys Ala
 115 120 125
 Leu Lys Thr His Ser Gly Ser Cys Arg Gln Cys Met Arg Leu Ser Lys
 130 135 140
 10 Cys Gly Pro Gly Phe Gly Val Ala Ser Ser Arg Ala Pro Asn Gly Asn
 145 150 155 160
 Val Leu Cys Lys Ala Cys Ala Pro Gly Thr Phe Ser Asp Thr Thr Ser
 165 170 175
 15 Ser Thr Asp Val Cys Arg Pro His Arg Ile Cys Ser Ile Leu Ala Ile
 180 185 190
 20 Pro Gly Asn Ala Ser Thr Asp Ala Val Cys Ala Pro Glu Ser Pro Thr
 195 200 205
 Leu Ser Ala Ile Pro Arg Thr Leu Tyr Val Ser Gln Pro Glu Pro Thr
 210 215 220
 25 Arg Ser Gln
 225
 30 <210> 14
 <211> 225
 <212> PRT
 <213> Unknown
 35 <220>
 <223> Description of Unknown Organism: primate
 <400> 14
 40 Met Ala Pro Val Ala Val Trp Ala Ala Leu Ala Val Gly Leu Glu Leu
 1 5 10 15
 Trp Ala Ala Ala His Ala Leu Pro Ala Gln Val Ala Phe Thr Pro Tyr
 20 25 30
 45 Ala Pro Glu Pro Gly Ser Thr Cys Arg Leu Arg Glu Tyr Tyr Asp Gln
 35 40 45
 Thr Ala Gln Met Cys Cys Ser Lys Cys Ser Pro Gly Gln His Ala Lys
 50 55 60
 50 Val Phe Cys Thr Lys Thr Ser Asp Thr Val Cys Asp Ser Cys Glu Asp
 65 70 75 80
 Ser Thr Tyr Thr Gln Leu Trp Asn Trp Val Pro Glu Cys Leu Ser Cys
 85 90 95
 55 Gly Ser Arg Cys Ser Ser Asp Gln Val Glu Thr Gln Ala Cys Thr Arg
 100 105 110
 60 Glu Gln Asn Arg Ile Cys Thr Cys Arg Pro Gly Trp Tyr Cys Ala Leu
 115 120 125
 Ser Lys Gln Glu Gly Cys Arg Leu Cys Ala Pro Leu Arg Lys Cys Arg
 130 135 140

SUBSTITUTE SHEET (rule 26)

5 Pro Gly Phe Gly Val Ala Arg Pro Gly Thr Glu Thr Ser Asp Val Val
 145 150 155 160
 Cys Lys Pro Cys Ala Pro Gly Thr Phe Ser Asn Thr Thr Ser Ser Thr
 165 170 175
 10 Asp Ile Cys Arg Pro His Gln Ile Cys Asn Val Val Ala Ile Pro Gly
 180 185 190
 15 Asn Ala Ser Met Asp Ala Val Cys Thr Ser Thr Ser Pro Thr Arg Ser
 195 200 205
 Met Ala Pro Gly Ala Val His Leu Pro Gln Pro Val Ser Thr Arg Ser
 210 215 220
 20 Gln
 225
 <210> 15
 25 <211> 187
 <212> PRT
 <213> Unknown
 <220>
 30 <223> Description of Unknown Organism:primate
 <400> 15
 Met Asn Lys Leu Leu Cys Cys Ala Leu Val Phe Leu Asp Ile Ser Ile
 1 5 10 15
 35 Lys Trp Thr Thr Gln Glu Thr Phe Pro Pro Lys Tyr Leu His Tyr Asp
 20 25 30
 40 Glu Glu Thr Ser His Gln Leu Leu Cys Asp Lys Cys Pro Pro Gly Thr
 35 40 45
 Tyr Leu Lys Gln His Cys Thr Ala Lys Trp Lys Thr Val Cys Ala Pro
 50 55 60
 45 Cys Pro Asp His Tyr Tyr Thr Asp Ser Trp His Thr Ser Asp Glu Cys
 65 70 75 80
 Leu Tyr Cys Ser Pro Val Cys Lys Glu Leu Gln Tyr Val Lys Gln Glu
 85 90 95
 50 Cys Asn Arg Thr His Asn Arg Val Cys Glu Cys Lys Glu Gly Arg Tyr
 100 105 110
 Leu Glu Ile Glu Phe Cys Leu Lys His Arg Ser Cys Pro Pro Gly Phe
 115 120 125
 55 Gly Val Val Gln Ala Gly Thr Pro Glu Arg Asn Thr Val Cys Lys Arg
 130 135 140
 60 Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser Lys Ala Pro Cys
 145 150 155 160
 Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu Leu Thr Gln Lys
 165 170 175

SUBSTITUTE SHEET (rule 26)

5 Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly
180 185

10 <210> 16
<211> 636
<212> DNA
<213> Unknown

15 <220>
<223> Description of Unknown Organism: rodent

<220>
<221> CDS
<222> (104)..(553)

20 <400> 16
ggcacgaggg cgtttgccgc ggaagtgcta ccaagctgcg gaaagcgtga gtctggagca 60

25 cagcactggc gactagcagg aataaacacg tttggtgaga gcc atg gca ctc aag 115
Met Ala Leu Lys
1

30 gtc cta cct cta cac agg acg gtg ctc ttc gct gcc att ctc ttc cta 163
Val Leu Pro Leu His Arg Thr Val Leu Phe Ala Ala Ile Leu Phe Leu
5 10 15 20

35 ctc cac ctg gca tgt aaa gtg agt tgc gaa acc gga gat tgc agg cag 211
Leu His Leu Ala Cys Lys Val Ser Cys Glu Thr Gly Asp Cys Arg Gln
25 30 35

40 cag gaa ttc aag gat cga tct gga aac tgt gtc ctc tgc aaa cag tgc 259
Gln Glu Phe Lys Asp Arg Ser Gly Asn Cys Val Leu Cys Lys Gln Cys
40 45 50

45 gga cct ggc atg gag ttg tcc aag gaa tgt ggc ttc ggc tat ggg gag 307
Gly Pro Gly Met Glu Leu Ser Lys Glu Cys Gly Phe Gly Tyr Gly Glu
55 60 65

50 gat gca cag tgt gtg ccc tgc agg ccg cac cgg ttc aag gaa gac tgg 355
Asp Ala Gln Cys Val Pro Cys Arg Pro His Arg Phe Lys Glu Asp Trp
70 75 80

55 ggt ttc cag aag tgt aag cca tgt gcg gac tgt gcg ctg gtg aac cgc 403
Gly Phe Gln Lys Cys Lys Pro Cys Ala Asp Cys Ala Leu Val Asn Arg
85 90 95 100

60 ttt cag agg gcc aac tgc tca cac acc agt gat gct gtc tgc ggg gac 451
Phe Gln Arg Ala Asn Cys Ser His Thr Ser Asp Ala Val Cys Gly Asp
105 110 115

65 tgc ctg cca gga ttt tac cgg aag acc aaa ctg gtt ggt ttt caa gac 499
Cys Leu Pro Gly Phe Tyr Arg Lys Thr Lys Leu Val Gly Phe Gln Asp
120 125 130

70 atg gag tgt gtg ccc tgc gga gac cca cct cct ccc tac gaa cca cac 547
Met Glu Cys Val Pro Cys Gly Asp Pro Pro Pro Pro Tyr Glu Pro His
135 140 145

30

5 tgt gag tgatgtgcc aagtggcagca gacctttaaa aaaaaaagaa aaaaaaacia 603
 Cys Glu
 150

 acaaaaacia aaaaaaaaaa aaaaaaaaaa aaa 636

10 <210> 17
 <211> 150
 <212> PRT
 <213> Unknown

15 <400> 17
 Met Ala Leu Lys Val Leu Pro Leu His Arg Thr Val Leu Phe Ala Ala
 1 5 10 15

20 Ile Leu Phe Leu Leu His Leu Ala Cys Lys Val Ser Cys Glu Thr Gly
 20 25 30

 Asp Cys Arg Gln Gln Glu Phe Lys Asp Arg Ser Gly Asn Cys Val Leu
 35 40 45

25 Cys Lys Gln Cys Gly Pro Gly Met Glu Leu Ser Lys Glu Cys Gly Phe
 50 55 60

30 Gly Tyr Gly Glu Asp Ala Gln Cys Val Pro Cys Arg Pro His Arg Phe
 65 70 75 80

 Lys Glu Asp Trp Gly Phe Gln Lys Cys Lys Pro Cys Ala Asp Cys Ala
 85 90 95

35 Leu Val Asn Arg Phe Gln Arg Ala Asn Cys Ser His Thr Ser Asp Ala
 100 105 110

 Val Cys Gly Asp Cys Leu Pro Gly Phe Tyr Arg Lys Thr Lys Leu Val
 115 120 125

40 Gly Phe Gln Asp Met Glu Cys Val Pro Cys Gly Asp Pro Pro Pro Pro
 130 135 140

45 Tyr Glu Pro His Cys Glu
 145 150

50 <210> 18
 <211> 474
 <212> DNA
 <213> Unknown

55 <220>
 <223> Description of Unknown Organism:primate

 <220>
 <221> CDS
 <222> (78)..(473)

60 <400> 18

cgcgctgagg tggatttgta ccggagtccc atttgggagc aagagccatc tactcgctccg 60
 ttaccggcct tcccacc atg gat tgc caa gaa aat gag tac tgg gac caa 110

SUBSTITUTE SHEET (rule 26)

31

	Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln																
5	1				5				10								
	tgg gga cgg tgt gtc acc tgc caa cgg tgt ggt cct gga cag gag cta	15				20				25				158			
	Trp Gly Arg Cys Val Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu																
10	tcc aag gat tgt ggt tat gga gag ggt gga gat gcc tac tgc aca gcc	30				35				40				206			
	Ser Lys Asp Cys Gly Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala																
15	tgc cct cct cgc agt aca aaa gca gct ggg gcc acc aca aat gtc aga	45				50				55				254			
	Cys Pro Pro Arg Ser Thr Lys Ala Ala Gly Ala Thr Asn Val Arg																
20	gtt gca tca cct gtg ctg tca tca atc gtg ttc aga agg ttc aac tgc	60				65				70				302			
	Val Ala Ser Pro Val Leu Ser Ser Ile Val Phe Arg Arg Phe Asn Cys																
25	aca gtn acc tct nat gct gtc tgt ggg gga ngg ttt gcc caa gtt tct	80				85				90				350			
	Thr Xaa Thr Ser Xaa Ala Val Cys Gly Gly Xaa Phe Ala Gln Val Ser																
30	aac cga aag aca cgc cat tgg aag gct gcc agg acc aag gat ggc atc	95				100				105				398			
	Asn Arg Lys Thr Arg His Trp Lys Ala Ala Arg Thr Lys Asp Gly Ile																
35	ccg tgg cac aaa gnc aga ccc cca act tct gan ggt tnc aaa gtg nct	110				115				120				446			
	Pro Trp His Lys Xaa Arg Pro Pro Thr Ser Xaa Gly Xaa Lys Val Xaa																
40	ttc caa ttg gag ctt aat ggg agg can a	125				130								474			
	Phe Gln Leu Glu Leu Asn Gly Arg Xaa																
45	<210> 19																
	<211> 132																
	<212> PRT																
	<213> Unknown																
50	<400> 19																
	Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln Trp Gly Arg Cys Val	1				5				10				15			
	Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu Ser Lys Asp Cys Gly	20				25				30							
	Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala Cys Pro Pro Arg Ser	35				40				45							
55	Thr Lys Ala Ala Gly Ala Thr Thr Asn Val Arg Val Ala Ser Pro Val	50				55				60							
	Leu Ser Ser Ile Val Phe Arg Arg Phe Asn Cys Thr Xaa Thr Ser Xaa	65				70				75				80			
60	Ala Val Cys Gly Gly Xaa Phe Ala Gln Val Ser Asn Arg Lys Thr Arg	85				90				95							

SUBSTITUTE SHEET (rule 26)

32

5 His Trp Lys Ala Ala Arg Thr Lys Asp Gly Ile Pro Trp His Lys Xaa
 100 105 110
 Arg Pro Pro Thr Ser Xaa Gly Xaa Lys Val Xaa Phe Gln Leu Glu Leu
 115 120 125
 10 Asn Gly Arg Xaa
 130
 15 <210> 20
 <211> 546
 <212> DNA
 <213> Unknown
 20 <220>
 <223> Description of Unknown Organism: primate
 <220>
 <221> CDS
 <222> (78)..(308)
 25 <400> 20
 cgcgctgagg tggatttgta ccggagtccc atttgggagc aagagccatc tactcgtccg 60
 30 ttaccggcct tcccacc atg gat tgc caa gaa aat gag tac tgg gac caa 110
 Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln
 1 5 10
 tgg gga cgg tgt gtc acc tgc caa cgg tgt ggt cct gga cag gag cta 158
 Trp Gly Arg Cys Val Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu
 15 20 25
 35 tcc aag gat tgt ggt tat gga gag ggt gga gat gcc tac tgc aca gcc 206
 Ser Lys Asp Cys Gly Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala
 30 35 40
 40 tgc cct cct cgc agg tac aaa agc agc tgg ggc cac cac aaa tgt cag 254
 Cys Pro Pro Arg Arg Tyr Lys Ser Ser Trp Gly His His Lys Cys Gln
 45 50 55
 45 agt tgc atc acc tgt gct gtc atc aat cgt gtt cag aag gtc caa ctg 302
 Ser Cys Ile Thr Cys Ala Val Ile Asn Arg Val Gln Lys Val Gln Leu
 60 65 70 75
 50 cac agc taacctctna tgctgtctgt ggggatgttt gncccaagtt ctnaccgaaa 358
 His Ser
 agacacgcca tgggaaggct ggcaggacca ngaatggccn tcccgtggca gaaagccaga 418
 55 ccccccaacn nctgnagggt ccaatgtggc cttgccattt ggaagcttan tgggaaggca 478
 gatgncaacc caaagtggcc ccttcaggga ggccaaaatt tggtggcaat gggtnagca 538
 gcntgcca 546

SUBSTITUTE SHEET (rule 26)

33

5 <210> 21
 <211> 77
 <212> PRT
 <213> Unknown
 <400> 21

10 Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln Trp Gly Arg Cys Val
 1 5 10 15

15 Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu Ser Lys Asp Cys Gly
 20 25 30

Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala Cys Pro Pro Arg Arg
 35 40 45

20 Tyr Lys Ser Ser Trp Gly His His Lys Cys Gln Ser Cys Ile Thr Cys
 50 55 60

Ala Val Ile Asn Arg Val Gln Lys Val Gln Leu His Ser
 65 70 75

25 <210> 22
 <211> 932
 <212> DNA
 <213> Unknown

30 <220>
 <223> Description of Unknown Organism:primate

35 <220>
 <221> CDS
 <222> (78)..(770)

40 <220>
 <221> misc_feature
 <222> (782)
 <223> nucleotide may be A, C, G, or T

<400> 22

45 cgcgctgagg tggatttgta ccggagtccc atttgggagc aagagccatc tactcgtccg 60
 ttaccggcct tcccacc atg gat tgc caa gaa aat gag tac tgg gac caa 110
 Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln
 1 5 10

50 tgg gga cgg tgt gtc acc tgc caa cgg tgt ggt cct gga cag gag cta 158
 Trp Gly Arg Cys Val Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu
 15 20 25

55 tcc aag gat tgt ggt tat gga gag ggt gga gat gcc tac tgc aca gcc 206
 Ser Lys Asp Cys Gly Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala
 30 35 40

60 tgc cct cct cgc agg tac aaa agc agc tgg ggc cac cac aaa tgt cag 254
 Cys Pro Pro Arg Arg Tyr Lys Ser Ser Trp Gly His His Lys Cys Gln
 45 50 55

SUBSTITUTE SHEET (rule 26)

34

5	agt tgc atc acc tgt gct gtc atc aat cgt gtt cag aag gtc aac tgc	302
	Ser Cys Ile Thr Cys Ala Val Ile Asn Arg Val Gln Lys Val Asn Cys	
	60 65 70 75	
10	aca gct acc tct aat gct gtc tgt ggg gac tgt ttg ccc agg ttc tac	350
	Thr Ala Thr Ser Asn Ala Val Cys Gly Asp Cys Leu Pro Arg Phe Tyr	
	80 85 90	
15	cga aag aca cgc att gga ggc ctg cag gac caa gag tgc atc ccg tgc	398
	Arg Lys Thr Arg Ile Gly Gly Leu Gln Asp Gln Glu Cys Ile Pro Cys	
	95 100 105	
20	acg aag cag acc ccc acc tct gag gtt caa tgt gcc ttc cag ttg agc	446
	Thr Lys Gln Thr Pro Thr Ser Glu Val Gln Cys Ala Phe Gln Leu Ser	
	110 115 120	
25	tta gtg gag gca gat gca ccc aca gtg ccc cct cag gag gcc aca ctt	494
	Leu Val Glu Ala Asp Ala Pro Thr Val Pro Pro Gln Glu Ala Thr Leu	
	125 130 135	
30	gtt gca ctg gtg agc agc ctg cta gtg gtg ttt acc ctg gcc ttc ctg	542
	Val Ala Leu Val Ser Ser Leu Leu Val Val Phe Thr Leu Ala Phe Leu	
	140 145 150 155	
35	ggg ctc ttc ttc ctc tac tgc aag cag ttc ttc aac aga cat tgc cag	590
	Gly Leu Phe Phe Leu Tyr Cys Lys Gln Phe Phe Asn Arg His Cys Gln	
	160 165 170	
40	cgt gga ggt ttg ctg cag ttt gag gct gat aaa aca gca aag gag gaa	638
	Arg Gly Gly Leu Leu Gln Phe Glu Ala Asp Lys Thr Ala Lys Glu Glu	
	175 180 185	
45	tct ctc ttc ccc gtg cca ccc agc aag gag acc agt gct gag tcc caa	686
	Ser Leu Phe Pro Val Pro Pro Ser Lys Glu Thr Ser Ala Glu Ser Gln	
	190 195 200	
50	gtc tct tgg gcc cct ggc agc ctt gcc cag ttg ttc tct ctg gac tct	734
	Val Ser Trp Ala Pro Gly Ser Leu Ala Gln Leu Phe Ser Leu Asp Ser	
	205 210 215	
55	gtt cct ata cca caa cag cag cag ggg cct gaa atg tgatgtccac	780
	Val Pro Ile Pro Gln Gln Gln Gln Gly Pro Glu Met	
	220 225 230	
60	angagctaataccctacagatggggcatatcctatcccatcccaccagag gattgattct	840
	ccatttcacaaggactgatacaggagcatttcttgcttccc tggtagtagtc tggggagcca	900
	gattccacat tcatgggact accagacatg tt	932
55	<210> 23	
	<211> 231	
	<212> PRT	
	<213> Unknown	
60	<400> 23	

Met Asp Cys Gln Glu Asn Glu Tyr Trp Asp Gln Trp Gly Arg Cys Val
1 5 10 15

SUBSTITUTE SHEET (rule 26)

35

5 Thr Cys Gln Arg Cys Gly Pro Gly Gln Glu Leu Ser Lys Asp Cys Gly
 20 25 30
 Tyr Gly Glu Gly Gly Asp Ala Tyr Cys Thr Ala Cys Pro Pro Arg Arg
 35 40 45
 10 Tyr Lys Ser Ser Trp Gly His His Lys Cys Gln Ser Cys Ile Thr Cys
 50 55 60
 Ala Val Ile Asn Arg Val Gln Lys Val Asn Cys Thr Ala Thr Ser Asn
 65 70 75 80
 Ala Val Cys Gly Asp Cys Leu Pro Arg Phe Tyr Arg Lys Thr Arg Ile
 85 90 95
 20 Gly Gly Leu Gln Asp Gln Glu Cys Ile Pro Cys Thr Lys Gln Thr Pro
 100 105 110
 Thr Ser Glu Val Gln Cys Ala Phe Gln Leu Ser Leu Val Glu Ala Asp
 115 120 125
 25 Ala Pro Thr Val Pro Pro Gln Glu Ala Thr Leu Val Ala Leu Val Ser
 130 135 140
 Ser Leu Leu Val Val Phe Thr Leu Ala Phe Leu Gly Leu Phe Phe Leu
 145 150 155 160
 Tyr Cys Lys Gln Phe Phe Asn Arg His Cys Gln Arg Gly Gly Leu Leu
 165 170 175
 35 Gln Phe Glu Ala Asp Lys Thr Ala Lys Glu Glu Ser Leu Phe Pro Val
 180 185 190
 Pro Pro Ser Lys Glu Thr Ser Ala Glu Ser Gln Val Ser Trp Ala Pro
 195 200 205
 40 Gly Ser Leu Ala Gln Leu Phe Ser Leu Asp Ser Val Pro Ile Pro Gln
 210 215 220
 Gln Gln Gln Gly Pro Glu Met
 45 225 230

 <210> 24
 <211> 232
 50 <212> DNA
 <213> Unknown

 <220>
 <223> Description of Unknown Organism: primate
 55
 <400> 24
 ttggccttag ggaccaagct tttatcatcg tcagtgggac ttaacctgtc ttaaaagtgc 60
 tgcttctcct acactcgctc aagatcccga gtcagctgta ttatggcatc ctattagtca 120
 60 ggcagcctgt gcttcaagcc cgtagttgta ttcacccct aaaggggcca ttccgtttgt 180
 atcatcacat gtcctcagtg ggtccatgtg tatatcaagg acatgatgca ga 232

SUBSTITUTE SHEET (rule 26)

36
 <210> 25
 <211> 77
 <212> PRT
 5 <213> Unknown
 <220>
 10 <223> Description of Unknown Organism:primate
 <400> 25
 Leu Ala Leu Gly Thr Lys Leu Leu Ser Ser Ser Val Gly Leu Asn Leu
 1 5 10 15
 15 Ser Xaa Lys Cys Cys Phe Ser Tyr Thr Arg Ser Arg Ser Arg Val Ser
 20 25 30
 Cys Ile Met Ala Ser Tyr Xaa Ser Gly Ser Leu Cys Phe Lys Pro Val
 20 35 40 45
 Val Val Phe Ile Pro Xaa Arg Gly His Ser Val Cys Ile Ile Thr Cys
 50 55 60
 25 Pro Gln Trp Val His Val Tyr Ile Lys Asp Met Met Gln
 65 70 75
 30 <210> 26
 <211> 72
 <212> PRT
 <213> Unknown
 <220>
 35 <223> Description of Unknown Organism:primate
 <400> 26
 Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro Ser Glu Cys
 1 5 10 15
 40 Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg Ile Met Asp
 20 25 30
 Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile Val Phe Ile
 45 35 40 45
 Thr Xaa Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp Lys Trp Val
 50 55 60
 50 Gln Asp Tyr Ile Lys Asp Met Lys
 65 70
 55 <210> 27
 <211> 143
 <212> PRT
 <213> Unknown
 <220>
 60 <223> Description of Unknown Organism:primate
 <400> 27
 Met Lys Ile Ser Val Ala Ala Ile Pro Phe Phe Leu Leu Ile Thr Ile
 1 5 10 15

SUBSTITUTE SHEET (rule 26)

5 Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro
20 25 30

Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg
35 40 45

10 Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile
50 55 60

Val Phe Ile Thr Lys Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp
65 70 75 80

15 Lys Trp Val Gln Asp Tyr Ile Lys Asp Met Lys Glu Asn Thr Lys Thr
85 90 95

20 Glu Ser Ser Ser Arg Gly Pro Tyr His Pro Ser Glu Cys Cys Phe Thr
100 105 110

Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg Ile Met Asp Tyr Tyr Glu
115 120 125

25 Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile Val Phe Ile Thr Xaa
130 135 140

<210> 28
<211> 93
<212> PRT
<213> Unknown

<220>
35 <223> Description of Unknown Organism:primate

<400> 28
Met Lys Ile Ser Val Ala Ala Ile Pro Phe Phe Leu Leu Ile Thr Ile
1 5 10 15

40 Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro
20 25 30

Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg
35 40 45

45 Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile
50 55 60

50 Val Phe Ile Thr Lys Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp
65 70 75 80

Lys Trp Val Gln Asp Tyr Ile Lys Asp Met Lys Glu Asn
85 90

55 <210> 29
<211> 93
<212> PRT
60 <213> Unknown

<220>
<223> Description of Unknown Organism:primate
<400> 29

SUBSTITUTE SHEET (rule 26)

38

5 Met Lys Ile Ser Val Ala Ala Ile Pro Phe Phe Leu Leu Ile Thr Ile
 1 5 10 15
 Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro
 20 25 30
 10 Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg
 35 40 45
 Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile
 50 55 60
 15 Val Phe Ile Thr Lys Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp
 65 70 75 80
 20 Lys Trp Val Gln Asp Tyr Ile Lys Asp Met Lys Glu Asn
 85 90
 <210> 30
 <211> 93
 25 <212> PRT
 <213> Unknown
 <220>
 <223> Description of Unknown Organism:primate
 30 <400> 30
 Met Lys Ile Ser Val Ala Ala Ile Pro Phe Phe Leu Leu Ile Thr Ile
 1 5 10 15
 35 Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro
 20 25 30
 Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg
 35 40 45
 40 Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile
 50 55 60
 Val Phe Ile Thr Lys Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp
 45 65 70 75 80
 Lys Trp Val Gln Asp Tyr Ile Lys Asp Met Lys Glu Asn
 85 90
 50 <210> 31
 <211> 1082
 <212> DNA
 <213> Unknown
 55 <220>
 <223> Description of Unknown Organism:primate
 <220>
 60 <221> CDS
 <222> (1)...(1080)
 <220>
 <221> misc_feature

SUBSTITUTE SHEET (rule 26)

5 <222> (20)
 <223> nucleotide may be G

 <220>
 <221> misc_feature
 <222> (56)
 10 <223> nucleotide may be A, C, G or T

 <220>
 <221> misc_feature
 <222> (103)
 15 <223> nucleotide may be A, C, G, or T

 <220>
 <221> misc_feature
 <222> (130)
 20 <223> nucleotide may be C or T

 <220>
 <221> misc_feature
 <222> (190)
 25 <223> nucleotide may be A or C

 <220>
 <221> misc_feature
 <222> (256)
 30 <223> nucleotide may be C or G

 <400> 31
 atg cct ttc ccc ggc cca cac gca ggt aga tct tcc act cta aag gac 48
 Met Pro Phe Pro Gly Pro His Ala Gly Arg Ser Ser Thr Leu Lys Asp
 35 1 5 10 15

 acc acc cct cca tcc cac caa ata ttt gga agg ctc ctg gaa gat ctc 96
 Thr Thr Pro Pro Ser His Gln Ile Phe Gly Arg Leu Leu Glu Asp Leu
 20 25 30
 40
 caa atc caa gtg tct ccc act gcc cac ggc att cca gac act ttt gac 144
 Gln Ile Gln Val Ser Pro Thr Ala His Gly Ile Pro Asp Thr Phe Asp
 35 40 45
 45 cct tac ctg gac atc gcc ctg gat atc cag gca gct cag agt gtc cag 192
 Pro Tyr Leu Asp Ile Ala Leu Asp Ile Gln Ala Ala Gln Ser Val Gln
 50 55 60
 50 caa gct ttg gaa cag ttg gtg aag ccc gaa gaa ctc aat gga gag aat 240
 Gln Ala Leu Glu Gln Leu Val Lys Pro Glu Glu Leu Asn Gly Glu Asn
 65 70 75 80
 gcc tat cat tgt ggt ctt tgt ctc cag agg gcg ccg gcc tcc aag acg 288
 Ala Tyr His Cys Gly Leu Cys Leu Gln Arg Ala Pro Ala Ser Lys Thr
 55 85 90 95
 tta act tta cac acc tct gcc aag gtc ctc atc ctt gtc ttg aag aga 336
 Leu Thr Leu His Thr Ser Ala Lys Val Leu Ile Leu Val Leu Lys Arg
 100 105 110
 60 ttc tcc gat gtc aca ggc aac aag att gcc aag aat gtg caa tat cct 384
 Phe Ser Asp Val Thr Gly Asn Lys Ile Ala Lys Asn Val Gln Tyr Pro
 115 120 125

40

5	gag tgc ctt gac atg cag cca tac atg tct cag cag aac aca gga cct Glu Cys Leu Asp Met Gln Pro Tyr Met Ser Gln Gln Asn Thr Gly Pro 130 135 140	432
10	ctt gtc tat gtc ctc tat gct gtg ctg gtc cac gct ggg tgg agt tgt Leu Val Tyr Val Leu Tyr Ala Val Leu Val His Ala Gly Trp Ser Cys 145 150 155 160	480
15	cac aac gga cat tac ttc tct tat gtc aaa gct caa gaa ggc cag tgg His Asn Gly His Tyr Phe Ser Tyr Val Lys Ala Gln Glu Gly Gln Trp 165 170 175	528
20	tat aaa atg gat gat gcc gag gtc acc gcc tct agc atc act tct gtc Tyr Lys Met Asp Asp Ala Glu Val Thr Ala Ser Ser Ile Thr Ser Val 180 185 190	576
25	ctg agt caa cag gcc tac gtc ctc ttt tac atc cag aag agt gaa tgg Leu Ser Gln Gln Ala Tyr Val Leu Phe Tyr Ile Gln Lys Ser Glu Trp 195 200 205	624
30	gaa aga cac agt gag agt gtg tca aga ggc agg gaa cca aga gcc ctt Glu Arg His Ser Glu Ser Val Ser Arg Gly Arg Glu Pro Arg Ala Leu 210 215 220	672
35	ggc gca gaa gac aca gac agg cga gca acg caa gga gag ctc aag aga Gly Ala Glu Asp Thr Asp Arg Arg Ala Thr Gln Gly Glu Leu Lys Arg 225 230 235 240	720
40	gac cac ccc tgc ctc cag gcc ccc gag ttg gac gag cac ttg gtg gaa Asp His Pro Cys Leu Gln Ala Pro Glu Leu Asp Glu His Leu Val Glu 245 250 255	768
45	aga gcc act cag gaa agc acc tta gac cac tgg aaa ttc ctt caa gag Arg Ala Thr Gln Glu Ser Thr Leu Asp His Trp Lys Phe Leu Gln Glu 260 265 270	816
50	caa aac aaa acg aag cct gag ttc aac gtc aga aaa gtc gaa ggt acc Gln Asn Lys Thr Lys Pro Glu Phe Asn Val Arg Lys Val Glu Gly Thr 275 280 285	864
55	ctg cct ccc gac gta ctt gtg att cat caa tca aaa tac aag tgt ggg Leu Pro Pro Asp Val Leu Val Ile His Gln Ser Lys Tyr Lys Cys Gly 290 295 300	912
60	atg aag aac cat cat cct gaa cag caa agc tcc ctg cta aac ctc tct Met Lys Asn His His Pro Glu Gln Gln Ser Ser Leu Leu Asn Leu Ser 305 310 315 320	960
65	tcg acg acc ccg aca cat cag gag tcc atg aac act ggc aca ctc gct Ser Thr Thr Pro Thr His Gln Glu Ser Met Asn Thr Gly Thr Leu Ala 325 330 335	1008
70	tcc ctg cga ggg agg gcc agg aga tcc aaa ggg aag aac aaa cac agc Ser Leu Arg Gly Arg Ala Arg Arg Ser Lys Gly Lys Asn Lys His Ser 340 345 350	1056
75	aag agg gct ctg ctt gtg tgc cag tg Lys Arg Ala Leu Leu Val Cys Gln 355 360	1082

SUBSTITUTE SHEET (rule 26)

41

5 <210> 32
 <211> 360
 <212> PRT
 <213> Unknown

 <400> 32
 10 Met Pro Phe Pro Gly Pro His Ala Gly Arg Ser Ser Thr Leu Lys Asp
 1 5 10 15

 Thr Thr Pro Pro Ser His Gln Ile Phe Gly Arg Leu Leu Glu Asp Leu
 20 25 30
 15 Gln Ile Gln Val Ser Pro Thr Ala His Gly Ile Pro Asp Thr Phe Asp
 35 40 45

 Pro Tyr Leu Asp Ile Ala Leu Asp Ile Gln Ala Ala Gln Ser Val Gln
 50 55 60
 20 Gln Ala Leu Glu Gln Leu Val Lys Pro Glu Glu Leu Asn Gly Glu Asn
 65 70 75 80

 Ala Tyr His Cys Gly Leu Cys Leu Gln Arg Ala Pro Ala Ser Lys Thr
 85 90 95
 25 Leu Thr Leu His Thr Ser Ala Lys Val Leu Ile Leu Val Leu Lys Arg
 100 105 110
 30 Phe Ser Asp Val Thr Gly Asn Lys Ile Ala Lys Asn Val Gln Tyr Pro
 115 120 125

 Glu Cys Leu Asp Met Gln Pro Tyr Met Ser Gln Gln Asn Thr Gly Pro
 130 135 140
 35 Leu Val Tyr Val Leu Tyr Ala Val Leu Val His Ala Gly Trp Ser Cys
 145 150 155 160

 His Asn Gly His Tyr Phe Ser Tyr Val Lys Ala Gln Glu Gly Gln Trp
 165 170 175
 40 Tyr Lys Met Asp Asp Ala Glu Val Thr Ala Ser Ser Ile Thr Ser Val
 180 185 190
 45 Leu Ser Gln Gln Ala Tyr Val Leu Phe Tyr Ile Gln Lys Ser Glu Trp
 195 200 205

 Glu Arg His Ser Glu Ser Val Ser Arg Gly Arg Glu Pro Arg Ala Leu
 210 215 220
 50 Gly Ala Glu Asp Thr Asp Arg Arg Ala Thr Gln Gly Glu Leu Lys Arg
 225 230 235 240

 Asp His Pro Cys Leu Gln Ala Pro Glu Leu Asp Glu His Leu Val Glu
 245 250 255
 55 Arg Ala Thr Gln Glu Ser Thr Leu Asp His Trp Lys Phe Leu Gln Glu
 260 265 270
 60 Gln Asn Lys Thr Lys Pro Glu Phe Asn Val Arg Lys Val Glu Gly Thr
 275 280 285

SUBSTITUTE SHEET (rule 26)

42

5	Leu	Pro	Pro	Asp	Val	Leu	Val	Ile	His	Gln	Ser	Lys	Tyr	Lys	Cys	Gly	
	290						295					300					
	Met	Lys	Asn	His	His	Pro	Glu	Gln	Gln	Ser	Ser	Leu	Leu	Asn	Leu	Ser	
	305					310					315					320	
10	Ser	Thr	Thr	Pro	Thr	His	Gln	Glu	Ser	Met	Asn	Thr	Gly	Thr	Leu	Ala	
					325					330					335		
	Ser	Leu	Arg	Gly	Arg	Ala	Arg	Arg	Ser	Lys	Gly	Lys	Asn	Lys	His	Ser	
				340					345					350			
15	Lys	Arg	Ala	Leu	Leu	Val	Cys	Gln									
			355					360									
20	<210>	33															
	<211>	1683															
	<212>	DNA															
	<213>	Unknown															
25	<220>																
	<223>	Description of Unknown Organism:primate															
	<220>																
	<221>	CDS															
30	<222>	(1)..(1590)															
	<400>	33															
	atg	gag	gac	gac	tca	ctc	tac	ttg	gga	ggt	gag	tgg	cag	ttc	aac	cac	48
	Met	Glu	Asp	Asp	Ser	Leu	Tyr	Leu	Gly	Gly	Glu	Trp	Gln	Phe	Asn	His	
35	1				5					10					15		
	ttt	tca	aaa	ctc	aca	tct	tct	cgg	cca	gat	gca	gct	ttt	gct	gaa	atc	96
	Phe	Ser	Lys	Leu	Thr	Ser	Ser	Arg	Pro	Asp	Ala	Ala	Phe	Ala	Glu	Ile	
				20					25					30			
40	cag	cgg	act	tct	ctc	cct	gag	aag	tca	cca	ctc	tca	tct	gag	gcc	cgt	144
	Gln	Arg	Thr	Ser	Leu	Pro	Glu	Lys	Ser	Pro	Leu	Ser	Ser	Glu	Ala	Arg	
			35					40					45				
45	gtc	gac	ctc	tgt	gat	gat	ttg	gct	cct	gtg	gca	aga	cag	ctt	gct	ccc	192
	Val	Asp	Leu	Cys	Asp	Asp	Leu	Ala	Pro	Val	Ala	Arg	Gln	Leu	Ala	Pro	
		50					55					60					
50	agg	gag	aag	ctt	cct	ctg	agt	agc	agg	aga	cct	gct	gcg	gtg	ggg	gct	240
	Arg	Glu	Lys	Leu	Pro	Leu	Ser	Ser	Arg	Arg	Pro	Ala	Ala	Val	Gly	Ala	
		65				70					75				80		
	ggg	ctc	cag	aat	atg	gga	aat	acc	tgc	tac	gag	aac	gct	tcc	ctg	cag	288
	Gly	Leu	Gln	Asn	Met	Gly	Asn	Thr	Cys	Tyr	Glu	Asn	Ala	Ser	Leu	Gln	
55					85					90					95		
	tgc	ctg	aca	tac	aca	ccg	ccc	ctt	gcc	aac	t						

SUBSTITUTE SHEET (rule 26)

43

5	caa gct cac atc aca tgg gcc ctc cac agt cct ggt cat gtc atc cag	432
	Gln Ala His Ile Thr Trp Ala Leu His Ser Pro Gly His Val Ile Gln	
	130 135 140	
10	ccc tca cag gca ttg gct gct ggc ttc cat aga ggc aag cag gaa gat	480
	Pro Ser Gln Ala Leu Ala Ala Gly Phe His Arg Gly Lys Gln Glu Asp	
	145 150 155 160	
15	gcc cat gaa ttt ctc atg ttc act gtg gat gcc atg aaa aag gca tgc	528
	Ala His Glu Phe Leu Met Phe Thr Val Asp Ala Met Lys Lys Ala Cys	
	165 170 175	
20	ctt ccc ggc cac aag cag gta gat cat cac tct aag gac acc acc ctc	576
	Leu Pro Gly His Lys Gln Val Asp His His Ser Lys Asp Thr Thr Leu	
	180 185 190	
25	atc cac caa ata ttt gga ggc tgc tgg aga tct caa atc aag tgt ctc	624
	Ile His Gln Ile Phe Gly Gly Cys Trp Arg Ser Gln Ile Lys Cys Leu	
	195 200 205	
30	cac tgc cac ggg att cca gac act ttt gac cct tac ctg gac atc gcc	672
	His Cys His Gly Ile Pro Asp Thr Phe Asp Pro Tyr Leu Asp Ile Ala	
	210 215 220	
35	ctg gat atc cag gca gct cag agt gtc aag caa gct ttg gaa cag ttg	720
	Leu Asp Ile Gln Ala Ala Gln Ser Val Lys Gln Ala Leu Glu Gln Leu	
	225 230 235 240	
40	gtg aag ccc gaa gaa ctc aat gga gag aat gcc tat cat tgt ggt ctt	768
	Val Lys Pro Glu Leu Asn Gly Glu Asn Ala Tyr His Cys Gly Leu	
	245 250 255	
45	tgt ctc cag agg gcg ccg gcc tcc aag acg tta act tta cac act tct	816
	Cys Leu Gln Arg Ala Pro Ala Ser Lys Thr Leu Thr Leu His Thr Ser	
	260 265 270	
50	gcc aag gtc ctc atc ctt gtm ttg aag aga ttc tcc gat gtc aca ggc	864
	Ala Lys Val Leu Ile Leu Xaa Leu Lys Arg Phe Ser Asp Val Thr Gly	
	275 280 285	
55	aac aaa ctt gcc aag aat gtg caa tat cct gag tgc ctt gac atg cag	912
	Asn Lys Leu Ala Lys Asn Val Gln Tyr Pro Glu Cys Leu Asp Met Gln	
	290 295 300	
60	cca tac atg tct cag cag aac aca gga cct ctt gtc tat gtc ctc tat	960
	Pro Tyr Met Ser Gln Gln Asn Thr Gly Pro Leu Val Tyr Val Leu Tyr	
	305 310 315 320	
65	gct gtg ctg gtc cac gct ggg tgg agt tgt cac aac gga cat tac ttc	1008
	Ala Val Leu Val His Ala Gly Trp Ser Cys His Asn Gly His Tyr Phe	
	325 330 335	
70	tct tat gtc aaa gct caa gaa ggc cag tgg tat aaa atg gat gat gcc	1056
	Ser Tyr Val Lys Ala Gln Glu Gly Gln Trp Tyr Lys Met Asp Asp Ala	
	340 345 350	
75	gag gtc acc gcc tct agc atc act tct gtc ctg agt caa cag gcc tac	1104
	Glu Val Thr Ala Ser Ser Ile Thr Ser Val Leu Ser Gln Gln Ala Tyr	
	355 360 365	

SUBSTITUTE SHEET (rule 26)

5	gtc ctc ttt tac atc cag aag agt gaa tgg gaa aga cac agt gag agt Val Leu Phe Tyr Ile Gln Lys Ser Glu Trp Glu Arg His Ser Glu Ser 370 375 380	1152
10	gtg tca aga ggc agg gaa cca aga gcc ctt ggc gca gaa gac aca gac Val Ser Arg Gly Arg Glu Pro Arg Ala Leu Gly Ala Glu Asp Thr Asp 385 390 395 400	1200
15	agg cga gca acg caa gga gag ctc aag aga gac cac ccc tgc ctc cag Arg Arg Ala Thr Gln Gly Glu Leu Lys Arg Asp His Pro Cys Leu Gln 405 410 415	1248
20	gcc ccc gag ttg gac gag cac ttg gtg gaa aga gcc act cag gaa agc Ala Pro Glu Leu Asp Glu His Leu Val Glu Arg Ala Thr Gln Glu Ser 420 425 430	1296
25	acc tta gac cac tgg aaa ttc ctt caa gag caa aac aaa acg aag cct Thr Leu Asp His Trp Lys Phe Leu Gln Glu Gln Asn Lys Thr Lys Pro 435 440 445	1344
30	gag ttc aac gtc aga aaa gtc gaa ggt acc ctg cct ccc gac gta ctt Glu Phe Asn Val Arg Lys Val Glu Gly Thr Leu Pro Pro Asp Val Leu 450 455 460	1392
35	gtg att cat caa tca aaa tac aag tgt ggg atg aag aac cat cat cct Val Ile His Gln Ser Lys Tyr Lys Cys Gly Met Lys Asn His His Pro 465 470 475 480	1440
40	gaa cag caa agc tcc ctg cta aac ctc tct tgc acg acc ccg aca cat Glu Gln Gln Ser Ser Leu Leu Asn Leu Ser Ser Thr Thr Pro Thr His 485 490 495	1488
45	cag gag tcc atg aac act ggc aca ctc gct tcc ctg cga ggg agg gcc Gln Glu Ser Met Asn Thr Gly Thr Leu Ala Ser Leu Arg Gly Arg Ala 500 505 510	1536
50	agg aga tcc aaa ggg aag aac aaa cac agc aag agg gct ctg ctt gtg Arg Arg Ser Lys Gly Lys Asn Lys His Ser Lys Arg Ala Leu Leu Val 515 520 525	1584
55	tgc cag tgatctcagt ggaagtaccg acccacacgt aggggtgcac acacacacgc Cys Gln 530	1640
60	acacacacag acacacacat aactacaccc agaagcgcgc tga 540	1683
65	<210> 34 <211> 530 <212> PRT <213> Unknown	
70	<400> 34 Met Glu Asp Asp Ser Leu Tyr Leu Gly Gly Glu Trp Gln Phe Asn His 1 5 10 15	
75	Phe Ser Lys Leu Thr Ser Ser Arg Pro Asp Ala Ala Phe Ala Glu Ile 20 25 30	
80	Gln Arg Thr Ser Leu Pro Glu Lys Ser Pro Leu Ser Ser Glu Ala Arg 35 40 45	

SUBSTITUTE SHEET (rule 26)

45

5	Val	Asp	Leu	Cys	Asp	Asp	Leu	Ala	Pro	Val	Ala	Arg	Gln	Leu	Ala	Pro	50	55	60
	Arg	Glu	Lys	Leu	Pro	Leu	Ser	Ser	Arg	Arg	Pro	Ala	Ala	Val	Gly	Ala	65	70	75
10	Gly	Leu	Gln	Asn	Met	Gly	Asn	Thr	Cys	Tyr	Glu	Asn	Ala	Ser	Leu	Gln	85	90	95
	Cys	Leu	Thr	Tyr	Thr	Pro	Pro	Leu	Ala	Asn	Tyr	Met	Leu	Ser	Arg	Glu	100	105	110
15	His	Ser	Gln	Thr	Cys	Gln	Arg	Pro	Lys	Cys	Cys	Met	Leu	Cys	Thr	Met	115	120	125
	Gln	Ala	His	Ile	Thr	Trp	Ala	Leu	His	Ser	Pro	Gly	His	Val	Ile	Gln	130	135	140
20	Pro	Ser	Gln	Ala	Leu	Ala	Ala	Gly	Phe	His	Arg	Gly	Lys	Gln	Glu	Asp	145	150	155
	Ala	His	Glu	Phe	Leu	Met	Phe	Thr	Val	Asp	Ala	Met	Lys	Lys	Ala	Cys	165	170	175
25	Leu	Pro	Gly	His	Lys	Gln	Val	Asp	His	Ser	Lys	Asp	Thr	Thr	Leu		180	185	190
30	Ile	His	Gln	Ile	Phe	Gly	Gly	Cys	Trp	Arg	Ser	Gln	Ile	Lys	Cys	Leu	195	200	205
	His	Cys	His	Gly	Ile	Pro	Asp	Thr	Phe	Asp	Pro	Tyr	Leu	Asp	Ile	Ala	210	215	220
35	Leu	Asp	Ile	Gln	Ala	Ala	Gln	Ser	Val	Lys	Gln	Ala	Leu	Glu	Gln	Leu	225	230	235
	Val	Lys	Pro	Glu	Glu	Leu	Asn	Gly	Glu	Asn	Ala	Tyr	His	Cys	Gly	Leu	245	250	255
40	Cys	Leu	Gln	Arg	Ala	Pro	Ala	Ser	Lys	Thr	Leu	Thr	Leu	His	Thr	Ser	260	265	270
45	Ala	Lys	Val	Leu	Ile	Leu	Xaa	Leu	Lys	Arg	Phe	Ser	Asp	Val	Thr	Gly	275	280	285
	Asn	Lys	Leu	Ala	Lys	Asn	Val	Gln	Tyr	Pro	Glu	Cys	Leu	Asp	Met	Gln	290	295	300
50	Pro	Tyr	Met	Ser	Gln	Gln	Asn	Thr	Gly	Pro	Leu	Val	Tyr	Val	Leu	Tyr	305	310	315
	Ala	Val	Leu	Val	His	Ala	Gly	Trp	Ser	Cys	His	Asn	Gly	His	Tyr	Phe	325	330	335
55	Ser	Tyr	Val	Lys	Ala	Gln	Glu	Gly	Gln	Trp	Tyr	Lys	Met	Asp	Asp	Ala	340	345	350
60	Glu	Val	Thr	Ala	Ser	Ser	Ile	Thr	Ser	Val	Leu	Ser	Gln	Gln	Ala	Tyr	355	360	365

SUBSTITUTE SHEET (rule 26)

46

5 Val Leu Phe Tyr Ile Gln Lys Ser Glu Trp Glu Arg His Ser Glu Ser
 370 375 380
 Val Ser Arg Gly Arg Glu Pro Arg Ala Leu Gly Ala Glu Asp Thr Asp
 385 390 395 400
 10 Arg Arg Ala Thr Gln Gly Glu Leu Lys Arg Asp His Pro Cys Leu Gln
 405 410 415
 Ala Pro Glu Leu Asp Glu His Leu Val Glu Arg Ala Thr Gln Glu Ser
 420 425 430
 15 Thr Leu Asp His Trp Lys Phe Leu Gln Glu Gln Asn Lys Thr Lys Pro
 435 440 445
 Glu Phe Asn Val Arg Lys Val Glu Gly Thr Leu Pro Pro Asp Val Leu
 450 455 460
 Val Ile His Gln Ser Lys Tyr Lys Cys Gly Met Lys Asn His His Pro
 465 470 475 480
 25 Glu Gln Gln Ser Ser Leu Leu Asn Leu Ser Ser Thr Thr Pro Thr His
 485 490 495
 Gln Glu Ser Met Asn Thr Gly Thr Leu Ala Ser Leu Arg Gly Arg Ala
 500 505 510
 30 Arg Arg Ser Lys Gly Lys Asn Lys His Ser Lys Arg Ala Leu Leu Val
 515 520 525
 Cys Gln
 35 530
 <210> 35
 <211> 735
 40 <212> DNA
 <213> Unknown
 <220>
 <223> Description of Unknown Organism:primate
 45 <220>
 <221> CDS
 <222> (1)..(735)
 50 <220>
 <221> misc_feature
 <222> (197)
 <223> nucleotide may be A or C
 55 <220>
 <221> misc_feature
 <222> (590)
 <223> nucleotide may be A, C, G, or T
 60 <220>
 <221> misc_feature
 <222> (664)
 <223> nucleotide may be A, C, G, or T

SUBSTITUTE SHEET (rule 26)

47

5 <220>
 <221> misc_feature
 <222> (665)
 <223> nucleotide may be A, C, G, or T

10 <220>
 <221> misc_feature
 <222> (666)
 <223> nucleotide may be A, C, G, or T

15 <220>
 <221> misc_feature
 <222> (708)
 <223> nucleotide may be A, C, G, or T

20 <220>
 <221> misc_feature
 <222> (715)
 <223> nucleotide may be A, C, G, or T

25 <400> 35
 atg gct gtg cca agt tgg atc gtc aaa cgc agg cta cta cct tgg tcc 48
 Met Ala Val Pro Ser Trp Ile Val Lys Arg Arg Leu Leu Pro Trp Ser
 1 5 10 15

30 atc aaa ttt ttg gag ggt atc tca gat cac ggc gtg aag tgc tcc gtg 96
 Ile Lys Phe Leu Glu Gly Ile Ser Asp His Gly Val Lys Cys Ser Val
 20 25 30

35 tgc aag agc gtc tcg gac acc tac gac ccc tac ttg gac gtc gcg ctg 144
 Cys Lys Ser Val Ser Asp Thr Tyr Asp Pro Tyr Leu Asp Val Ala Leu
 35 40 45

40 gag atc cgg caa gct gcg aat att gtg cgt gct ctg gaa ctt ttt gtg 192
 Glu Ile Arg Gln Ala Ala Asn Ile Val Arg Ala Leu Glu Leu Phe Val
 50 55 60

45 aaa gca gat gtc ctg agt gga gag aat gcc tac atg tgt gct aaa tgc 240
 Lys Ala Asp Val Leu Ser Gly Glu Asn Ala Tyr Met Cys Ala Lys Cys
 65 70 75 80

50 aag aag aag gtt cca gcc agc aag cgc ttc acc atc cac aga aca tcc 288
 Lys Lys Lys Val Pro Ala Ser Lys Arg Phe Thr Ile His Arg Thr Ser
 85 90 95

55 aac gtc tta acc ctt tcc ctc aag cgc ttt gcc aac ttc agc ggg ggg 336
 Asn Val Leu Thr Leu Ser Leu Lys Arg Phe Ala Asn Phe Ser Gly Gly
 100 105 110

60 aag atc acc aag gat gta ggc tat ccg gaa ttc ctc aac ata cgt ccg 384
 Lys Ile Thr Lys Asp Val Gly Tyr Pro Glu Phe Leu Asn Ile Arg Pro
 115 120 125

65 tat atg tcc cag aat aat ggt gat cct gtc atg tat gga ctc tat gct 432
 Tyr Met Ser Gln Asn Asn Gly Asp Pro Val Met Tyr Gly Leu Tyr Ala
 130 135 140

70 gtc ctg gtg cac tcg ggc tac agc tgc cat gcc ggg cac tat tac tgc 480
 Val Leu Val His Ser Gly Tyr Ser Cys His Ala Gly His Tyr Tyr Cys
 145 150 155 160

SUBSTITUTE SHEET (rule 26)

5	tac	gtg	aag	gca	agc	aat	gga	cag	tgg	tac	cag	atg	aat	gat	tcc	ttg	528
	Tyr	Val	Lys	Ala	Ser	Asn	Gly	Gln	Trp	Tyr	Gln	Met	Asn	Asp	Ser	Leu	
					165					170					175		
10	gtc	cca	ttc	cag	caa	cgt	cca	agt	tgg	ttt	ctg	aaa	cca	gca	ggc	cta	576
	Val	Pro	Phe	Gln	Gln	Arg	Pro	Ser	Trp	Phe	Leu	Lys	Pro	Ala	Gly	Leu	
				180					185					190			
15	agt	ggc	ttg	ttc	tca	tcg	gcg	aat	ttc	cag	gct	ctc	aag	aaa	aat	tcc	624
	Ser	Gly	Leu	Phe	Ser	Ser	Ala	Asn	Phe	Gln	Ala	Leu	Lys	Lys	Asn	Ser	
			195					200					205				
20	cga	agg	gcc	tcc	att	ttc	cag	gaa	cag	gtt	cct	tcc	tcc	cct	tcc	cgg	672
	Arg	Arg	Ala	Ser	Ile	Phe	Gln	Glu	Gln	Val	Pro	Ser	Ser	Pro	Ser	Arg	
			210				215					220					
25	gcg	gcc	cga	att	gtg	aat	tcc	aga	ttc	att	ccc	agc	agg	aac	ctc	ggc	720
	Ala	Ala	Arg	Ile	Val	Asn	Ser	Arg	Phe	Ile	Pro	Ser	Arg	Asn	Leu	Gly	
			225			230					235					240	
30	aat	ggg	gat	tat	ttt												735
	Asn	Gly	Asp	Tyr	Phe												
					245												
35	<210>	36															
	<211>	245															
	<212>	PRT															
	<213>	Unknown															
40	<400>	36															
	Met	Ala	Val	Pro	Ser	Trp	Ile	Val	Lys	Arg	Arg	Leu	Leu	Pro	Trp	Ser	
	1				5					10					15		
45	Ile	Lys	Phe	Leu	Glu	Gly	Ile	Ser	Asp	His	Gly	Val	Lys	Cys	Ser	Val	
				20					25					30			
50	Cys	Lys	Ser	Val	Ser	Asp	Thr	Tyr	Asp	Pro	Tyr	Leu	Asp	Val	Ala	Leu	
			35					40					45				
55	Glu	Ile	Arg	Gln	Ala	Ala	Asn	Ile	Val	Arg	Ala	Leu	Glu	Leu	Phe	Val	
		50					55					60					
60	Lys	Ala	Asp	Val	Leu	Ser	Gly	Glu	Asn	Ala	Tyr	Met	Cys	Ala	Lys	Cys	
		65				70					75					80	
65	Lys	Lys	Lys	Val	Pro	Ala	Ser	Lys	Arg	Phe	Thr	Ile	His	Arg	Thr	Ser	
				85						90					95		
70	Asn	Val	Leu	Thr	Leu	Ser	Leu	Lys	Arg	Phe	Ala	Asn	Phe	Ser	Gly	Gly	
				100					105					110			
75	Lys	Ile	Thr	Lys	Asp	Val	Gly	Tyr	Pro	Glu	Phe	Leu	Asn	Ile	Arg	Pro	
			115					120					125				
80	Tyr	Met	Ser	Gln	Asn	Asn	Gly	Asp	Pro	Val	Met	Tyr	Gly				

5 Tyr Val Lys Ala Ser Asn Gly Gln Trp Tyr Gln Met Asn Asp Ser Leu
165 170 175

Val Pro Phe Gln Gln Arg Pro Ser Trp Phe Leu Lys Pro Ala Gly Leu
180 185 190

10 Ser Gly Leu Phe Ser Ser Ala Asn Phe Gln Ala Leu Lys Lys Asn Ser
195 200 205

Arg Arg Ala Ser Ile Phe Gln Glu Gln Val Pro Ser Ser Pro Ser Arg
210 215 220

15 Ala Ala Arg Ile Val Asn Ser Arg Phe Ile Pro Ser Arg Asn Leu Gly
225 230 235 240

20 Asn Gly Asp Tyr Phe
245

<210> 37
<211> 2244
25 <212> DNA
<213> Unknown

<220>
30 <223> Description of Unknown Organism:primate

<220>
<221> CDS
<222> (1)..(2244)

35 <400> 37
atg cag aaa gcc tgc ctg aat ggc tgt gcc aag ttg gat cgt caa acg 48
Met Gln Lys Ala Cys Leu Asn Gly Cys Ala Lys Leu Asp Arg Gln Thr
1 5 10 15

40 cag gct act acc ttg gtc cat caa att ttt gga ggg tat ctc aga tca 96
Gln Ala Thr Thr Leu Val His Gln Ile Phe Gly Gly Tyr Leu Arg Ser
20 25 30

45 cgc gtg aag tgc tcc gtg tgc aag agc gtc tcg gac acc tac gac ccc 144
Arg Val Lys Cys Ser Val Cys Lys Ser Val Ser Asp Thr Tyr Asp Pro
35 40 45

50 tac ttg gac gtc gcg ctg gag atc cgg caa gct gcg aat att gtg cgt 192
Tyr Leu Asp Val Ala Leu Glu Ile Arg Gln Ala Ala Asn Ile Val Arg
50 55 60

55 gct ctg gaa ctt ttt gtg aaa gca gat gtc ctg agt gga gag aat gcc 240
Ala Leu Glu Leu Phe Val Lys Ala Asp Val Leu Ser Gly Glu Asn Ala
65 70 75 80

60 tac atg tgt gct aaa tgc aag aag aag gtt cca gcc agc aag cgc ttc 288
Tyr Met Cys Ala Lys Cys Lys Lys Lys Val Pro Ala Ser Lys Arg Phe
85 90 95

acc atc cac aga aca tcc aac gtc tta acc ctt tcc ctc aag cgc ttt 336
Thr Ile His Arg Thr Ser Asn Val Leu Thr Leu Ser Leu Lys Arg Phe
100 105 110

gcc aac ttc agc ggg ggg aag atc acc aag gat gta ggc tat ccg gaa 384

SUBSTITUTE SHEET (rule 26)

50

5	Ala Asn Phe Ser Gly Gly Lys Ile Thr Lys Asp Val Gly Tyr Pro Glu	115	120	125	
	ttc ctc aac ata cgt ccg tat atg tcc cag aat aat ggt gat cct gtc				432
	Phe Leu Asn Ile Arg Pro Tyr Met Ser Gln Asn Asn Gly Asp Pro Val	130	135	140	
10	atg tat gga ctc tat gct gtc ctg gtg cac tcg ggc tac agc tgc cat				480
	Met Tyr Gly Leu Tyr Ala Val Leu Val His Ser Gly Tyr Ser Cys His	145	150	155	160
15	gcc ggg cac tat tac tgc tac gtg aag gca agc aat gga cag tgg tac				528
	Ala Gly His Tyr Tyr Cys Tyr Val Lys Ala Ser Asn Gly Gln Trp Tyr	165	170		175
20	cag atg aat gat tcc ttg gtc cat tcc agc aac gtc aag gtg gtt ctg				576
	Gln Met Asn Asp Ser Leu Val His Ser Ser Asn Val Lys Val Val Leu	180	185		190
25	aac cag cag gcc tac gtg ctg ttc tat ctg cga att cca ggc tct aag				624
	Asn Gln Gln Ala Tyr Val Leu Phe Tyr Leu Arg Ile Pro Gly Ser Lys	195	200	205	
30	aaa agt ccc gag ggc ctc atc tcc agg aca ggc tcc tcc tcc ctt ccc				672
	Lys Ser Pro Glu Gly Leu Ile Ser Arg Thr Gly Ser Ser Ser Leu Pro	210	215	220	
	ggc cgc ccg agt gtg att cca gat cac tcc aag aag aac atc ggc aat				720
	Gly Arg Pro Ser Val Ile Pro Asp His Ser Lys Lys Asn Ile Gly Asn	225	230	235	240
35	ggg att att tcc tcc cca ctg act gga aag cga caa gac tct ggg acg				768
	Gly Ile Ile Ser Ser Pro Leu Thr Gly Lys Arg Gln Asp Ser Gly Thr	245	250		255
40	atg aag aag ccg cac acc act gaa gag att ggt gtg ccc ata tcc agg				816
	Met Lys Lys Pro His Thr Thr Glu Glu Ile Gly Val Pro Ile Ser Arg	260	265		270
45	aat ggc tcc acc ctg ggc ctg aag tcc cag aac ggc tgc att cct cca				864
	Asn Gly Ser Thr Leu Gly Leu Lys Ser Gln Asn Gly Cys Ile Pro Pro	275	280	285	
50	aag ctg ccc tcg ggg tcc cct tcc ccc aaa ctc tcc cag aca ccc aca				912
	Lys Leu Pro Ser Gly Ser Pro Ser Pro Lys Leu Ser Gln Thr Pro Thr	290	295	300	
	cac atg cca acc atc cta gac gac cct gga aag aag gtg aag aag cca				960
	His Met Pro Thr Ile Leu Asp Asp Pro Gly Lys Lys Val Lys Lys Pro	305	310	315	320
55	gct cct cca cag cac ttt tcc ccc aga act gct cag ggg ctg cct ggg				1008
	Ala Pro Pro Gln His Phe Ser Pro Arg Thr Ala Gln Gly Leu Pro Gly	325	330	335	
60	acc agc aac tcg aat agc agc aga tct ggg agc caa agg cag ggc tcc				1056
	Thr Ser Asn Ser Asn Ser Ser Arg Ser Gly Ser Gln Arg Gln Gly Ser	340	345	350	
	tgg gac agc agg gat gtt gtc ctc tct acc tca cct aag ctc ctg gct				1104

SUBSTITUTE SHEET (rule 26)

5	Trp	Asp	Ser	Arg	Asp	Val	Val	Leu	Ser	Thr	Ser	Pro	Lys	Leu	Leu	Ala	
			355					360					365				
	aca	gcc	act	gcc	aac	ggg	cat	ggg	ctg	aag	ggg	aac	gac	gag	agc	gct	1152
	Thr	Ala	Thr	Ala	Asn	Gly	His	Gly	Leu	Lys	Gly	Asn	Asp	Glu	Ser	Ala	
			370				375					380					
10	ggc	ctc	gac	agg	agg	ggc	tcc	agc	agc	tcc	agc	cca	gag	cac	tcg	gcc	1200
	Gly	Leu	Asp	Arg	Arg	Gly	Ser	Ser	Ser	Ser	Ser	Pro	Glu	His	Ser	Ala	
						390					395					400	
15	agc	agc	gac	tcc	acc	aag	gcc	ccc	cag	acc	ccc	agg	agt	gga	gcg	gcc	1248
	Ser	Ser	Asp	Ser		405	Lys	Ala	Pro	Gln	Thr	Pro	Arg	Ser	Gly	Ala	
										410					415		
20	cat	ctc	tgc	gat	tct	cag	gaa	acg	aac	tgt	tcc	acc	gct	ggc	cac	tcc	1296
	His	Leu	Cys	Asp	Ser	Gln	Glu	Thr	Asn	Cys	Ser	Thr	Ala	Gly	His	Ser	
				420					425					430			
25	aaa	acg	ccg	cca	agt	gga	gca	gat	tct	aag	acg	gtg	aag	ctg	aag	tcc	1344
	Lys	Thr	Pro	Pro	Ser	Gly	Ala	Asp	Ser	Lys	Thr	Val	Lys	Leu	Lys	Ser	
			435					440					445				
30	cct	gtc	ctg	agc	aac	acc	acc	act	gag	cct	gca	agc	acc	atg	tct	cct	1392
	Pro	Val	Leu	Ser	Asn	Thr	Thr	Thr	Glu	Pro	Ala	Ser	Thr	Met	Ser	Pro	
			450				455					460					
35	cca	cca	gcc	aaa	aaa	ctg	gcc	ctt	tct	gcc	aag	aag	gcc	agc	acc	ctg	1440
	Pro	Pro	Ala	Lys	Lys	Leu	Ala	Leu	Ser	Ala	Lys	Lys	Ala	Ser	Thr	Leu	
			465			470				475						480	
40	tgg	agg	gcg	acc	ggc	aat	gac	ctc	cgt	cca	cct	ccc	ccc	tca	cca	tcc	1488
	Trp	Arg	Ala	Thr	Gly	Asn	Asp	Leu	Arg	Pro	Pro	Pro	Pro	Ser	Pro	Ser	
					485				490						495		
45	tcc	gac	ctc	acc	cac	ccc	atg	aaa	acc	tct	cac	ccc	gtc	gtt	gcc	tcc	1536
	Ser	Asp	Leu	Thr	His	Pro	Met	Lys	Thr	Ser	His	Pro	Val	Val	Ala	Ser	
				500				505					510				
50	act	tgg	ccc	gtc	cat	aga	gcc	agg	gct	gtg	tca	cct	gct	ccc	caa	tca	1584
	Thr	Trp	Pro	Val	His	Arg	Ala	Arg	Ala	Val	Ser	Pro	Ala	Pro	Gln	Ser	
			515				520					525					
55	tcc	agc	cgc	ctg	caa	ccc	ccc	ttc	agc	ccc	cac	ccc	aca	ttg	ctg	tcc	1632
	Ser	Ser	Arg	Leu	Gln	Pro	Pro	Phe	Ser	Pro	His	Pro	Thr	Leu	Leu	Ser	
			530			535					540						
60	agt	acc	ccc	aag	ccc	cca	ggg	acg	tca	gaa	cca	cgg	agc	tgc	tcc	tcc	1680
	Ser	Thr	Pro	Lys	Pro	Pro	Gly	Thr	Ser	Glu	Pro	Arg	Ser	Cys	Ser	Ser	
						550					555					560	
65	atc	tcg	acg	gcg	ctg	cct	cag	gtc	aac	gag	gac	ctt	gtg	tct	ctt	cca	1728
	Ile	Ser	Thr	Ala	Leu	Pro	Gln	Val	Asn	Glu	Asp	Leu	Val	Ser	Leu	Pro	
					565				570						575		
70	cac	cag	ttg	cca	gag	gcc	agt	gag	ccc	ccc	cag	agc	ccc	tct	gag	aag	1776
	His	Gln	Leu	Pro	Glu	Ala	Ser	Glu	Pro	Pro	Gln	Ser	Pro	Ser	Glu	Lys	
				580				585						590			

5	agg aaa aag acc ttt gtg gga gag ccg cag agg ctg ggc tca gag acg Arg Lys Lys Thr Phe Val Gly Glu Pro Gln Arg Leu Gly Ser Glu Thr 595 600 605	1824
10	cgc ctc cca cag cac atc agg gag gcc act gcg gct ccc cac ggg aag Arg Leu Pro Gln His Ile Arg Glu Ala Thr Ala Ala Pro His Gly Lys 610 615 620	1872
15	agg aag agg aag aag aag aag cgc ccg gag gac aca gct gcc agc gcc Arg Lys Arg Lys Lys Lys Lys Arg Pro Glu Asp Thr Ala Ala Ser Ala 625 630 635 640	1920
20	ctg cag gag ggg cag aca cag aga cag cct ggg agc ccc atg tac agg Leu Gln Glu Gly Gln Thr Gln Arg Gln Pro Gly Ser Pro Met Tyr Arg 645 650 655	1968
25	agg gag ggc cag gca cag ctg ccc gct gtc aga cgg cag gaa gat ggc Arg Glu Gly Gln Ala Gln Leu Pro Ala Val Arg Arg Gln Glu Asp Gly 660 665 670	2016
30	aca cag cca cag gtg aat ggc cag cag gtg gga tgt gtt acg gac ggc Thr Gln Pro Gln Val Asn Gly Gln Gln Val Gly Cys Val Thr Asp Gly 675 680 685	2064
35	cac cac gcg agc agc agg aag cgg agg agg aaa gga gca gaa ggt ctt His His Ala Ser Ser Arg Lys Arg Arg Arg Lys Gly Ala Glu Gly Leu 690 695 700	2112
40	ggt gaa gaa ggc ggc ctg cac cag gac cca ctt cgg cac agc tgc tct Gly Glu Glu Gly Gly Leu His Gln Asp Pro Leu Arg His Ser Cys Ser 705 710 715 720	2160
45	ccc atg ggt gat ggt gat cca gag gcc atg gaa gag tct cca agg aaa Pro Met Gly Asp Gly Asp Pro Glu Ala Met Glu Glu Ser Pro Arg Lys 725 730 735	2208
50	aag aaa aaa aaa aaa aac tcg agg ggg ggc ccg gta Lys Lys Lys Lys Lys Asn Ser Arg Gly Gly Pro Val 740 745	2244
55	<210> 38 <211> 748 <212> PRT <213> Unknown	
60	<400> 38 Met Gln Lys Ala Cys Leu Asn Gly Cys Ala Lys Leu Asp Arg Gln Thr 1 5 10 15 Gln Ala Thr Thr Leu Val His Gln Ile Phe Gly Gly Tyr Leu Arg Ser 20 25 30 Arg Val Lys Cys Ser Val Cys Lys Ser Val Ser Asp Thr Tyr Asp Pro 35 40 45 Tyr Leu Asp Val Ala Leu Glu Ile Arg Gln Ala Ala Asn Ile Val Arg 50 55 60 Ala Leu Glu Leu Phe Val Lys Ala Asp Val Leu Ser Gly Glu Asn Ala 65 70 75 80	

5 Tyr Met Cys Ala Lys Cys Lys Lys Lys Val Pro Ala Ser Lys Arg Phe
 85 90 95
 Thr Ile His Arg Thr Ser Asn Val Leu Thr Leu Ser Leu Lys Arg Phe
 100 105 110
 10 Ala Asn Phe Ser Gly Gly Lys Ile Thr Lys Asp Val Gly Tyr Pro Glu
 115 120 125
 15 Phe Leu Asn Ile Arg Pro Tyr Met Ser Gln Asn Asn Gly Asp Pro Val
 130 135 140
 Met Tyr Gly Leu Tyr Ala Val Leu Val His Ser Gly Tyr Ser Cys His
 145 150 155 160
 20 Ala Gly His Tyr Tyr Cys Tyr Val Lys Ala Ser Asn Gly Gln Trp Tyr
 165 170 175
 Gln Met Asn Asp Ser Leu Val His Ser Ser Asn Val Lys Val Val Leu
 180 185 190
 25 Asn Gln Gln Ala Tyr Val Leu Phe Tyr Leu Arg Ile Pro Gly Ser Lys
 195 200 205
 30 Lys Ser Pro Glu Gly Leu Ile Ser Arg Thr Gly Ser Ser Ser Leu Pro
 210 215 220
 Gly Arg Pro Ser Val Ile Pro Asp His Ser Lys Lys Asn Ile Gly Asn
 225 230 235 240
 35 Gly Ile Ile Ser Ser Pro Leu Thr Gly Lys Arg Gln Asp Ser Gly Thr
 245 250 255
 Met Lys Lys Pro His Thr Thr Glu Glu Ile Gly Val Pro Ile Ser Arg
 260 265 270
 40 Asn Gly Ser Thr Leu Gly Leu Lys Ser Gln Asn Gly Cys Ile Pro Pro
 275 280 285
 45 Lys Leu Pro Ser Gly Ser Pro Ser Pro Lys Leu Ser Gln Thr Pro Thr
 290 295 300
 His Met Pro Thr Ile Leu Asp Asp Pro Gly Lys Lys Val Lys Lys Pro
 305 310 315 320
 50 Ala Pro Pro Gln His Phe Ser Pro Arg Thr Ala Gln Gly Leu Pro Gly
 325 330 335
 Thr Ser Asn Ser Asn Ser Ser Arg Ser Gly Ser Gln Arg Gln Gly Ser
 340 345 350
 55 Trp Asp Ser Arg Asp Val Val Leu Ser Thr Ser Pro Lys Leu Leu Ala
 355 360 365
 60 Thr Ala Thr Ala Asn Gly His Gly Leu Lys Gly Asn Asp Glu Ser Ala
 370 375 380
 Gly Leu Asp Arg Arg Gly Ser Ser Ser Ser Ser Pro Glu His Ser Ala
 385 390 395 400

5 Ser Ser Asp Ser Thr Lys Ala Pro Gln Thr Pro Arg Ser Gly Ala Ala
 405 410 415
 His Leu Cys Asp Ser Gln Glu Thr Asn Cys Ser Thr Ala Gly His Ser
 420 425 430
 10 Lys Thr Pro Pro Ser Gly Ala Asp Ser Lys Thr Val Lys Leu Lys Ser
 435 440 445
 Pro Val Leu Ser Asn Thr Thr Thr Glu Pro Ala Ser Thr Met Ser Pro
 450 455 460
 15 Pro Pro Ala Lys Lys Leu Ala Leu Ser Ala Lys Lys Ala Ser Thr Leu
 465 470 475 480
 20 Trp Arg Ala Thr Gly Asn Asp Leu Arg Pro Pro Pro Pro Ser Pro Ser
 485 490 495
 Ser Asp Leu Thr His Pro Met Lys Thr Ser His Pro Val Val Ala Ser
 500 505 510
 25 Thr Trp Pro Val His Arg Ala Arg Ala Val Ser Pro Ala Pro Gln Ser
 515 520 525
 Ser Ser Arg Leu Gln Pro Pro Phe Ser Pro His Pro Thr Leu Leu Ser
 530 535 540
 30 Ser Thr Pro Lys Pro Pro Gly Thr Ser Glu Pro Arg Ser Cys Ser Ser
 545 550 555 560
 35 Ile Ser Thr Ala Leu Pro Gln Val Asn Glu Asp Leu Val Ser Leu Pro
 565 570 575
 His Gln Leu Pro Glu Ala Ser Glu Pro Pro Gln Ser Pro Ser Glu Lys
 580 585 590
 40 Arg Lys Lys Thr Phe Val Gly Glu Pro Gln Arg Leu Gly Ser Glu Thr
 595 600 605
 Arg Leu Pro Gln His Ile Arg Glu Ala Thr Ala Ala Pro His Gly Lys
 610 615 620
 45 Arg Lys Arg Lys Lys Lys Lys Arg Pro Glu Asp Thr Ala Ala Ser Ala
 625 630 635 640
 50 Leu Gln Glu Gly Gln Thr Gln Arg Gln Pro Gly Ser Pro Met Tyr Arg
 645 650 655
 Arg Glu Gly Gln Ala Gln Leu Pro Ala Val Arg Arg Gln Glu Asp Gly
 660 665 670
 55 Thr Gln Pro Gln Val Asn Gly Gln Gln Val Gly Cys Val Thr Asp Gly
 675 680 685
 His His Ala Ser Ser Arg Lys Arg Arg Arg Lys Gly Ala Glu Gly Leu
 690 695 700
 60 Gly Glu Glu Gly Gly Leu His Gln Asp Pro Leu Arg His Ser Cys Ser
 705 710 715 720

5	Pro Met Gly Asp Gly Asp Pro Glu Ala Met Glu Ser Pro Arg Lys	725	730	735
	Lys Lys Lys Lys Lys Asn Ser Arg Gly Gly Pro Val	740	745	
10	<210> 39			
	<211> 526			
	<212> PRT			
	<213> Unknown			
15	<220>			
	<223> Description of Unknown Organism:primate			
	<400> 39			
20	Met Val Val Ala Leu Ser Phe Pro Glu Ala Asp Pro Ala Leu Ser Ser	1	5	10
				15
	Pro Asp Ala Pro Glu Leu His Gln Asp Glu Ala Gln Val Val Glu Glu	20	25	30
25	Leu Thr Val Asn Gly Lys His Ser Leu Ser Trp Glu Ser Pro Gln Gly	35	40	45
	Pro Gly Cys Gly Leu Gln Asn Thr Gly Asn Ser Cys Tyr Leu Asn Ala	50	55	60
30	Ala Leu Gln Cys Leu Thr His Thr Pro Pro Leu Ala Asp Tyr Met Leu	65	70	75
	Ser Gln Glu His Ser Gln Thr Cys Cys Ser Pro Glu Gly Cys Lys Leu	85	90	95
	Cys Ala Met Glu Ala Leu Val Thr Gln Ser Leu Leu His Ser His Ser	100	105	110
40	Gly Asp Val Met Lys Pro Ser His Ile Leu Thr Ser Ala Phe His Lys	115	120	125
	His Gln Gln Glu Asp Ala His Glu Phe Leu Met Phe Thr Leu Glu Thr	130	135	140
45	Met His Glu Ser Cys Leu Gln Val His Arg Gln Ser Lys Pro Thr Ser	145	150	155
	Glu Asp Ser Ser Pro Ile His Asp Ile Phe Gly Gly Trp Trp Arg Ser	165	170	175
	Gln Ile Lys Cys Leu Leu Cys Gln Gly Thr Ser Asp Thr Tyr Asp Arg	180	185	190
55	Phe Leu Asp Ile Pro Leu Asp Ile Ser Ser Ala Gln Ser Val Lys Gln	195	200	205
	Ala Leu Trp Asp Thr Glu Lys Ser Glu Glu Leu Cys Gly Asp Asn Ala	210	215	220
60	Tyr Tyr Cys Gly Lys Cys Arg Gln Lys Met Pro Ala Ser Lys Thr Leu	225	230	235
				240

56

5 His Val His Ile Ala Pro Lys Val Leu Met Val Val Leu Asn Arg Phe
 245 250 255
 Ser Ala Phe Thr Gly Asn Lys Leu Asp Arg Lys Val Ser Tyr Pro Glu
 260 265 270
 10 Phe Leu Asp Leu Lys Pro Tyr Leu Ser Glu Pro Thr Gly Gly Pro Leu
 275 280 285
 Pro Tyr Ala Leu Tyr Ala Val Leu Val His Asp Gly Ala Thr Ser His
 290 295 300
 15 Ser Gly His Tyr Phe Cys Cys Val Lys Ala Gly His Gly Lys Trp Tyr
 305 310 315 320
 Lys Met Asp Asp Thr Lys Val Thr Arg Cys Asp Val Thr Ser Val Leu
 325 330 335
 20 Asn Glu Asn Ala Tyr Val Leu Phe Tyr Val Gln Gln Ala Asn Leu Lys
 340 345 350
 25 Gln Val Ser Ile Asp Met Pro Glu Gly Arg Ile Asn Glu Val Leu Asp
 355 360 365
 Pro Glu Tyr Gln Leu Lys Lys Ser Arg Arg Lys Lys His Lys Lys Lys
 370 375 380
 30 Ser Pro Phe Thr Glu Asp Leu Gly Glu Pro Cys Glu Asn Arg Asp Lys
 385 390 395 400
 Arg Ala Ile Lys Glu Thr Ser Leu Gly Lys Gly Lys Val Leu Gln Glu
 405 410 415
 35 Val Asn His Lys Lys Ala Gly Gln Lys His Gly Asn Thr Lys Leu Met
 420 425 430
 40 Pro Gln Lys Gln Asn His Gln Lys Ala Gly Gln Asn Leu Arg Asn Thr
 435 440 445
 Glu Val Glu Leu Asp Leu Pro Ala Asp Ala Ile Val Ile His Gln Pro
 450 455 460
 45 Arg Ser Thr Ala Asn Trp Gly Arg Asp Ser Pro Asp Lys Glu Asn Gln
 465 470 475 480
 Pro Leu His Asn Ala Asp Arg Leu Leu Thr Ser Gln Gly Pro Val Asn
 485 490 495
 50 Thr Trp Gln Leu Cys Arg Gln Glu Gly Arg Arg Arg Ser Lys Lys Gly
 500 505 510
 55 Gln Asn Lys Asn Lys Gln Gly Gln Arg Leu Leu Leu Val Cys
 515 520 525
 60 <210> 40
 <211> 545
 <212> PRT
 <213> Unknown
 <220>

SUBSTITUTE SHEET (rule 26)

57

5 <223> Description of Unknown Organism:primate
 <400> 40
 Met Val Val Ser Leu Ser Phe Pro Glu Ala Asp Pro Ala Leu Ser Ser
 1 5 10 15
 10 Pro Gly Ala Gln Gln Leu His Gln Asp Glu Ala Gln Val Val Val Glu
 20 25 30
 Leu Thr Ala Asn Asp Lys Pro Ser Leu Ser Trp Glu Cys Pro Gln Gly
 35 40 45
 15 Pro Gly Cys Gly Leu Gln Asn Thr Gly Asn Ser Cys Tyr Leu Asn Ala
 50 55 60
 20 Ala Leu Gln Cys Leu Thr His Thr Pro Pro Leu Ala Asp Tyr Met Leu
 65 70 75 80
 Ser Gln Glu Tyr Ser Gln Thr Cys Cys Ser Pro Glu Gly Cys Lys Met
 85 90 95
 25 Cys Ala Met Glu Ala His Val Thr Gln Ser Leu Leu His Ser His Ser
 100 105 110
 Gly Asp Val Met Lys Pro Ser Gln Ile Leu Thr Ser Ala Phe His Lys
 115 120 125
 30 His Gln Gln Glu Asp Ala His Glu Phe Leu Met Phe Thr Leu Glu Thr
 130 135 140
 35 Met His Glu Ser Cys Leu Gln Val His Arg Gln Ser Glu Pro Thr Ser
 145 150 155 160
 Glu Asp Ser Ser Pro Ile His Asp Ile Phe Gly Gly Leu Trp Arg Ser
 165 170 175
 40 Gln Ile Lys Cys Leu His Cys Gln Gly Thr Ser Asp Thr Tyr Asp Arg
 180 185 190
 Phe Leu Asp Val Pro Leu Asp Ile Ser Ser Ala Gln Ser Val Asn Gln
 195 200 205
 45 Ala Leu Trp Asp Thr Glu Lys Ser Glu Glu Leu Arg Gly Glu Asn Ala
 210 215 220
 50 Tyr Tyr Cys Gly Arg Cys Arg Gln Lys Met Pro Ala Ser Lys Thr Leu
 225 230 235 240
 His Ile His Ser Ala Pro Lys Val Leu Leu Leu Val Leu Lys Arg Phe
 245 250 255
 55 Ser Ala Phe Met Gly Asn Lys Leu Asp Arg Lys Val Ser Tyr Pro Glu
 260 265 270
 Phe Leu Asp Leu Lys Pro Tyr Leu Ser Gln Pro Thr Gly Gly Pro Leu
 275 280 285
 60 Pro Tyr Ala Leu Tyr Ala Val Leu Val His Glu Gly Ala Thr Cys His
 290 295 300

SUBSTITUTE SHEET (rule 26)

58

5 Ser Gly His Tyr Phe Ser Tyr Val Lys Ala Arg His Gly Ala Trp Tyr
 305 310 315 320
 Lys Met Asp Asp Thr Lys Val Thr Ser Cys Asp Val Thr Ser Val Leu
 325 330 335
 10 Asn Glu Asn Ala Tyr Val Leu Phe Tyr Val Gln Gln Thr Asp Leu Lys
 340 345 350
 Gln Val Ser Ile Asp Met Pro Glu Gly Arg Val His Glu Val Leu Asp
 355 360 365
 15 Pro Glu Tyr Gln Leu Lys Lys Ser Arg Arg Lys Lys His Lys Lys Lys
 370 375 380
 20 Ser Pro Cys Thr Glu Asp Ala Gly Glu Pro Cys Lys Asn Arg Glu Lys
 385 390 395 400
 Arg Ala Thr Lys Glu Thr Ser Leu Gly Glu Gly Lys Val Xaa Gln Glu
 405 410 415
 25 Lys Asn His Lys Lys Ala Gly Gln Lys His Glu Asn Thr Lys Leu Val
 420 425 430
 Pro Gln Glu Gln Asn His Gln Lys Leu Gly Gln Lys His Arg Ile Asn
 435 440 445
 30 Glu Ile Leu Pro Gln Glu Gln Asn His Gln Lys Ala Gly Gln Ser Leu
 450 455 460
 35 Arg Asn Thr Glu Gly Glu Leu Asp Leu Pro Ala Asp Ala Ile Val Ile
 465 470 475 480
 His Leu Leu Arg Ser Thr Glu Asn Trp Gly Arg Asp Ala Pro Asp Lys
 485 490 495
 40 Glu Asn Gln Pro Trp His Asn Ala Asp Arg Leu Leu Thr Ser Gln Asp
 500 505 510
 Pro Val Asn Thr Gly Gln Leu Cys Arg Gln Glu Gly Arg Arg Arg Ser
 515 520 525
 45 Lys Lys Gly Lys Asn Lys Asn Lys Gln Gly Gln Arg Leu Leu Leu Val
 530 535 540
 50 Cys
 545
 <210> 41
 <211> 890
 55 <212> DNA
 <213> Unknown
 <220>
 <223> Description of Unknown Organism: primate
 60 <220>
 <221> CDS
 <222> (15) .. (500)

SUBSTITUTE SHEET (rule 26)

60

5 tct gtt ttg aat ttc tcc tat ccc atc tgt gag gcg gct ctg ccc aag 338
 Ser Val Leu Asn Phe Ser Tyr Pro Ile Cys Glu Ala Ala Leu Pro Lys
 95 100 105

10 ttt tct ttc tgt gga aga agg aaa gga gag cag att tac tat gct ggg 386
 Phe Ser Phe Cys Gly Arg Arg Lys Gly Glu Gln Ile Tyr Tyr Ala Gly
 110 115 120

15 cct gtc aat aat cct gaa ttt act att cct cag gga gaa tac cag gtt 434
 Pro Val Asn Asn Pro Glu Phe Thr Ile Pro Gln Gly Glu Tyr Gln Val
 125 130 135 140

20 ttg ctg gaa ctg tac act gaa aaa cgg tcc acc gtg gcc tgt gcc aat 482
 Leu Leu Glu Leu Tyr Thr Glu Lys Arg Ser Thr Val Ala Cys Ala Asn
 145 150 155

25 gct act atc atg tgc tcc tgactgtggg cctgttagca aaaactcaca 530
 Ala Thr Ile Met Cys Ser
 160

30 gccagctgca tctcgtcggg aaccttccaa gctcctctga ctgaacctac tgtgggagga 590
 gaagcagctg atgacagaga gaggctctac aaagaagcgc ccccaaagag tgcagctgct 650
 aatttttagtc ccaggaccag acatccccag actccacaga tgtaatgaag tccccgaatg 710

35 tatctgtttc taaggagcct cttggcagtc cttaagcagt cttgagggtc catccttttt 770
 ctctaattgg tcgcctccca ccagactcac ctgctttttca acttttttagg agtgcttcct 830
 cacacgttac caataataaa gaaagctggc caccaaaaaa aaaaaaaaaa aaaaaaaaaa 890

40 <210> 42
 <211> 162
 <212> PRT
 <213> Unknown

45 <400> 42
 Met Lys Gly Phe Thr Ala Thr Leu Phe Leu Trp Thr Leu Ile Phe Pro
 1 5 10 15

50 Ser Cys Ser Gly Gly Gly Gly Lys Ala Trp Pro Thr His Val Val
 20 25 30

55 Cys Ser Asp Ser Arg Leu Glu Val Leu Tyr Gln Ser Cys Asp Pro Leu
 35 40 45

60 Gln Asp Phe Gly Phe Ser Val Glu Lys Cys Ser Lys Gln Leu Lys Ser
 50 55 60

65 Asn Ile Asn Ile Arg Phe Gly Ile Ile Leu Arg Glu Asp Ile Lys Glu
 65 70 75 80

70 Leu Phe Leu Asp Leu Ala Leu Met Ser Gln Gly Ser Ser Val Leu Asn
 85 90 95

75 Phe Ser Tyr Pro Ile Cys Glu Ala Ala Leu Pro Lys Phe Ser Phe Cys
 100 105 110

SUBSTITUTE SHEET (rule 26)

61

5 Gly Arg Arg Lys Gly Glu Gln Ile Tyr Tyr Ala Gly Pro Val Asn Asn
115 120 125

Pro Glu Phe Thr Ile Pro Gln Gly Glu Tyr Gln Val Leu Leu Glu Leu
130 135 140

10 Tyr Thr Glu Lys Arg Ser Thr Val Ala Cys Ala Asn Ala Thr Ile Met
145 150 155 160

Cys Ser

15

<210> 43
<211> 486
<212> DNA
20 <213> Unknown

<220>
<223> Description of Unknown Organism:primate

25 <220>
<221> CDS
<222> (1)..(132)

<400> 43

30 ccc ctg ttt tct tcc ata ttt act gaa gct cag aag cag tat tgg gtc 48
Pro Leu Phe Ser Ser Ile Phe Thr Glu Ala Gln Lys Gln Tyr Trp Val
1 5 10 15

tgc aac tca tcc gat gca agt att tca tac acc tac tgt gat aaa atg 96
35 Cys Asn Ser Ser Asp Ala Ser Ile Ser Tyr Thr Tyr Cys Asp Lys Met
20 25 30

caa tac cca att tca att aat gtt aac ccc tgt ata gaattgaaag 142
40 Gln Tyr Pro Ile Ser Ile Asn Val Asn Pro Cys Ile
35 40

gatccaaagg attattgcac attttctaca ttccaaggag agatttaaag caattatatt 202

45 tcaatctcta tataactgtc aacaccatga atcttccaaa gcgcaaagaa gttatttgcc 262

gaggatctga tgacgattac tctttttgca gagctctgaa gggagagact gtgaatacaa 322

caatatcatt ctccctcaag ggaataaaat tttctaaggg aaaatacaaaa tgtgttggtg 382

50 aagctatttc tgggagccca gaagaaatgc tcttttgctt ggagtttgctc atcctacacc 442

aacctaattc aaattagaat aaattgagta tttaaaaaaa aaaa 486

55 <210> 44
<211> 44
<212> PRT
<213> Unknown

60 <400> 44
Pro Leu Phe Ser Ser Ile Phe Thr Glu Ala Gln Lys Gln Tyr Trp Val
1 5 10 15

SUBSTITUTE SHEET (rule 26)

62

```

5   Cys Asn Ser Ser Asp Ala Ser Ile Ser Tyr Thr Tyr Cys Asp Lys Met
    20                      25                      30

    Gln Tyr Pro Ile Ser Ile Asn Val Asn Pro Cys Ile
    35                      40

10  <210> 45
    <211> 483
    <212> DNA
    <213> Unknown

15  <220>
    <223> Description of Unknown Organism:primate

    <220>
20  <221> CDS
    <222> (1)..(480)

    <400> 45
25  atg ttc cca ttt ctg ttt ttt tcc acc ctg ttt tct tcc ata ttt act 48
    Met Phe Pro Phe Leu Phe Phe Ser Thr Leu Phe Ser Ser Ile Phe Thr
    1                      5                      10                      15

    gaa gct cag aag cag tat tgg gtc tgc aac tca tcc gat gca agt att 96
    Glu Ala Gln Lys Gln Tyr Trp Val Cys Asn Ser Ser Asp Ala Ser Ile
    20                      25                      30

30  tca tac acc tac tgt gat aaa atg caa tac cca att tca att aat gtt 144
    Ser Tyr Thr Tyr Cys Asp Lys Met Gln Tyr Pro Ile Ser Ile Asn Val
    35                      40                      45

    aac ccc tgt ata gaa ttg aaa gga tcc aaa gga tta ttg cac att ttc 192
    Asn Pro Cys Ile Glu Leu Lys Gly Ser Lys Gly Leu Leu His Ile Phe
    50                      55                      60

40  tac att cca agg aga gat tta aag caa tta tat ttc aat ctc tat ata 240
    Tyr Ile Pro Arg Arg Asp Leu Lys Gln Leu Tyr Phe Asn Leu Tyr Ile
    65                      70                      75                      80

    act gtc aac acc atg aat ctt cca aag cgc aaa gaa gtt att tgc cga 288
    Thr Val Asn Thr Met Asn Leu Pro Lys Arg Lys Glu Val Ile Cys Arg
    85                      90                      95

50  gga tct gat gac gat tac tct ttt tgc aga gct ctg aag gga gag act 336
    Gly Ser Asp Asp Asp Tyr Ser Phe Cys Arg Ala Leu Lys Gly Glu Thr
    100                      105                      110

    gtg aat aca aca ata tca ttc tcc ttc aag gga ata aaa ttt tct aag 384
    Val Asn Thr Thr Ile Ser Phe Ser Phe Lys Gly Ile Lys Phe Ser Lys
    115                      120                      125

55  gga aaa tac aaa tgt gtt gtt gaa gct att tct ggg agc cca gaa gaa 432
    Gly Lys Tyr Lys Cys Val Val Glu Ala Ile Ser Gly Ser Pro Glu Glu
    130                      135                      140

60  atg ctc ttt tgc ttg gag ttt gtc atc cta cac caa cct aat tca aat 480
    Met Leu Phe Cys Leu Glu Phe Val Ile Leu His Gln Pro Asn Ser Asn
    145                      150                      155                      160

tag 483

```

SUBSTITUTE SHEET (rule 26)

63

5 <210> 46
 <211> 160
 <212> PRT
 <213> Unknown

10 <400> 46
 Met Phe Pro Phe Leu Phe Phe Ser Thr Leu Phe Ser Ser Ile Phe Thr
 1 5 10 15

15 Glu Ala Gln Lys Gln Tyr Trp Val Cys Asn Ser Ser Asp Ala Ser Ile
 20 25 30

Ser Tyr Thr Tyr Cys Asp Lys Met Gln Tyr Pro Ile Ser Ile Asn Val
 35 40 45

20 Asn Pro Cys Ile Glu Leu Lys Gly Ser Lys Gly Leu Leu His Ile Phe
 50 55 60

Tyr Ile Pro Arg Arg Asp Leu Lys Gln Leu Tyr Phe Asn Leu Tyr Ile
 65 70 75 80

25 Thr Val Asn Thr Met Asn Leu Pro Lys Arg Lys Glu Val Ile Cys Arg
 85 90 95

30 Gly Ser Asp Asp Asp Tyr Ser Phe Cys Arg Ala Leu Lys Gly Glu Thr
 100 105 110

Val Asn Thr Thr Ile Ser Phe Ser Phe Lys Gly Ile Lys Phe Ser Lys
 115 120 125

35 Gly Lys Tyr Lys Cys Val Val Glu Ala Ile Ser Gly Ser Pro Glu Glu
 130 135 140

Met Leu Phe Cys Leu Glu Phe Val Ile Leu His Gln Pro Asn Ser Asn
 145 150 155 160

40 <210> 47
 <211> 498
 <212> DNA
 <213> Unknown

45 <220>
 <223> Description of Unknown Organism:rodent

50 <220>
 <221> CDS
 <222> (53)..(394)

<400> 47

55 gtcgagtcgcg atggtcttcc tggcgagttt aaagtatcgg agatattaaa tc atg ttg 58
 Met Leu
 1

60 cca ttt att ctc ttt tcg acg ctg ctt tct ccc ata ttg act gaa tct 106
 Pro Phe Ile Leu Phe Ser Thr Leu Leu Ser Pro Ile Leu Thr Glu Ser
 5 10 15

gag aag caa cag tgg ttc tgc aac tcc tcc gat gca att att tcc tac 154

SUBSTITUTE SHEET (rule 26)

5	Glu Lys Gln Gln Trp Phe Cys Asn Ser Ser Asp Ala Ile Ile Ser Tyr 20 25 30	
	agt tat tgt gat cac ttg aaa ttc cct att tca att agt tct gaa ccc Ser Tyr Cys Asp His Leu Lys Phe Pro Ile Ser Ile Ser Ser Glu Pro 35 40 45 50	202
10	tgc ata aga ctg agg gga acc aat gga ttt gtg cat gtt gag ttc att Cys Ile Arg Leu Arg Gly Thr Asn Gly Phe Val His Val Glu Phe Ile 55 60 65	250
15	cca aga gga aac tta aaa tat tta tat ttc aac cta ttc atc agt gtc Pro Arg Gly Asn Leu Lys Tyr Leu Tyr Phe Asn Leu Phe Ile Ser Val 70 75 80	298
20	aac tcc ata gag ttg ccg aag cgt aag gaa gtt ctg tgc cat gga cat Asn Ser Ile Glu Leu Pro Lys Arg Lys Glu Val Leu Cys His Gly His 85 90 95	346
25	gat gat gac tat tct ttt tgc aga gct ctg aaa gga gga tat gct att Asp Asp Asp Tyr Ser Phe Cys Arg Ala Leu Lys Gly Gly Tyr Ala Ile 100 105 110	394
	tagaaaatat gagactgtga atacatcaat accattctct ttcgagggaa tactatttcc taagggccat tacagatgtg ttgcagaagc tattgctggg gata	454 498
30	<210> 48 <211> 114 <212> PRT <213> Unknown	
35	<400> 48 Met Leu Pro Phe Ile Leu Phe Ser Thr Leu Leu Ser Pro Ile Leu Thr 1 5 10 15	
40	Glu Ser Glu Lys Gln Gln Trp Phe Cys Asn Ser Ser Asp Ala Ile Ile 20 25 30	
45	Ser Tyr Ser Tyr Cys Asp His Leu Lys Phe Pro Ile Ser Ile Ser Ser 35 40 45	
	Glu Pro Cys Ile Arg Leu Arg Gly Thr Asn Gly Phe Val His Val Glu 50 55 60	
50	Phe Ile Pro Arg Gly Asn Leu Lys Tyr Leu Tyr Phe Asn Leu Phe Ile 65 70 75 80	
	Ser Val Asn Ser Ile Glu Leu Pro Lys Arg Lys Glu Val Leu Cys His 85 90 95	
55	Gly His Asp Asp Asp Tyr Ser Phe Cys Arg Ala Leu Lys Gly Gly Tyr 100 105 110	
	Ala Ile	
60	<210> 49 <211> 141	

65

5 <212> PRT
 <213> Unknown
 <220>
 <223> Description of Unknown Organism:primate
 10 <400> 49
 Met Leu Pro Phe Ile Leu Phe Ser Thr Leu Leu Ser Pro Ile Leu Thr
 1 5 10 15
 Glu Ser Glu Lys Gln Gln Trp Phe Cys Asn Ser Ser Asp Ala Ile Ile
 15 20 25 30
 Ser Tyr Ser Tyr Cys Asp His Leu Lys Phe Pro Ile Ser Ile Ser Ser
 35 40 45
 20 Glu Pro Cys Ile Arg Leu Arg Gly Thr Asn Gly Phe Val His Val Glu
 50 55 60
 Phe Ile Pro Arg Gly Asn Leu Lys Tyr Leu Tyr Phe Asn Leu Phe Ile
 65 70 75 80
 25 Ser Val Asn Ser Ile Glu Leu Pro Lys Arg Lys Glu Val Leu Cys His
 85 90 95
 30 Gly His Asp Asp Asp Tyr Ser Phe Cys Arg Ala Leu Lys Gly Glu Thr
 100 105 110
 Val Asn Thr Ser Ile Pro Phe Ser Phe Glu Gly Ile Leu Phe Pro Lys
 115 120 125
 35 Gly His Tyr Arg Cys Val Ala Glu Ala Ile Ala Gly Asp
 130 135 140
 <210> 50
 40 <211> 162
 <212> PRT
 <213> Unknown
 <220>
 45 <223> Description of Unknown Organism:rodent
 <400> 50
 Met Asn Gly Val Ala Ala Ala Leu Leu Val Trp Ile Leu Thr Ser Pro
 1 5 10 15
 50 Ser Ser Ser Asp His Gly Ser Glu Asn Gly Trp Pro Lys His Thr Ala
 20 25 30
 Cys Asn Ser Gly Gly Leu Glu Val Val Tyr Gln Ser Cys Asp Pro Leu
 55 35 40 45
 Gln Asp Phe Gly Leu Ser Ile Asp Gln Cys Ser Lys Gln Ile Gln Ser
 50 55 60
 60 Asn Leu Asn Ile Arg Phe Gly Ile Ile Leu Arg Gln Asp Ile Arg Lys
 65 70 75 80
 Leu Phe Leu Asp Ile Thr Leu Met Ala Lys Gly Ser Ser Ile Leu Asn
 85 90 95

SUBSTITUTE SHEET (rule 26)

66

5 Tyr Ser Tyr Pro Leu Cys Glu Glu Asp Gln Pro Lys Phe Ser Phe Cys
 100 105 110
 Gly Arg Arg Lys Gly Glu Gln Ile Tyr Tyr Ala Gly Pro Val Asn Asn
 115 120 125
 10 Pro Gly Leu Asp Val Pro Gln Gly Glu Tyr Gln Leu Leu Leu Glu Leu
 130 135 140
 15 Tyr Asn Glu Asn Arg Ala Thr Val Ala Cys Ala Asn Ala Thr Val Thr
 145 150 155 160
 Ser Ser

 20 <210> 51
 <211> 1158
 <212> DNA
 <213> Unknown
 25 <220>
 <223> Description of Unknown Organism:avian

 <220>
 30 <221> CDS
 <222> (11)..(490)

 <400> 51
 35 cggtgcaacc atg aag aca ttg aat gtt ctc gct ctc gtc tta gtc ctg 49
 Met Lys Thr Leu Asn Val Leu Ala Leu Val Leu Val Leu
 1 5 10

 ctt tgc atc aat gcc agc aca gag tgg cct aca cac aca gtc tgc aag 97
 Leu Cys Ile Asn Ala Ser Thr Glu Trp Pro Thr His Thr Val Cys Lys
 15 20 25

 gag gaa aac ttg gag ata tat tac aaa agc tgt gat ccc cag caa gac 145
 Glu Glu Asn Leu Glu Ile Tyr Tyr Lys Ser Cys Asp Pro Gln Gln Asp
 30 35 40 45
 45 ttt gct ttc agc att gac cgt tgt tca gat gtc aca acc cac acc ttt 193
 Phe Ala Phe Ser Ile Asp Arg Cys Ser Asp Val Thr Thr His Thr Phe
 50 55 60

 50 gac atc aga gct gca atg gtc cta aga caa agc atc aag gaa ctg tat 241
 Asp Ile Arg Ala Ala Met Val Leu Arg Gln Ser Ile Lys Glu Leu Tyr
 65 70 75

 gcc aag gtt gat ctg atc ata aat ggg aag act gtc tta agc tac tca 289
 Ala Lys Val Asp Leu Ile Ile Asn Gly Lys Thr Val Leu Ser Tyr Ser
 80 85 90

 gag aca ctc tgt gga cca ggc ctt tct aag cta att ttc tgt gga aag 337
 Glu Thr Leu Cys Gly Pro Gly Leu Ser Lys Leu Ile Phe Cys Gly Lys
 95 100 105
 60 aag aaa gga gaa cat ctc tac tat gag gga cca atc aca ctg gga atc 385
 Lys Lys Gly Glu His Leu Tyr Tyr Glu Gly Pro Ile Thr Leu Gly Ile
 110 115 120 125

SUBSTITUTE SHEET (rule 26)

68

5 Cys Gly Pro Gly Leu Ser Lys Leu Ile Phe Cys Gly Lys Lys Lys Gly
 100 105 110
 Glu His Leu Tyr Tyr Glu Gly Pro Ile Thr Leu Gly Ile Lys Glu Ile
 115 120 125
 10 Pro Gln Arg Asp Tyr Thr Ile Thr Ala Arg Leu Thr Asn Glu Asp Arg
 130 135 140
 Ala Thr Val Ala Cys Ala Asp Phe Thr Val Lys Asn Tyr Leu Asp Tyr
 145 150 155 160
 15
 <210> 53
 <211> 2684
 <212> DNA
 20 <213> Unknown
 <220>
 <223> Description of Unknown Organism: primate
 25 <220>
 <221> CDS
 <222> (45)..(1256)
 <400> 53
 30 gaattcggca cgagggcgag gttttataca cctgaaagaa gaga atg tca aga cga 56
 Met Ser Arg Arg
 1
 35 agt agc cgt tta caa gct aag cag cag ccc cag ccc agc cag acg gaa 104
 Ser Ser Arg Leu Gln Ala Lys Gln Gln Pro Gln Pro Ser Gln Thr Glu 20
 5 10 15 20
 tcc ccc caa gaa gcc cag ata atc cag gcc aag aag agg aaa act acc 152
 Ser Pro Gln Glu Ala Gln Ile Ile Gln Ala Lys Lys Arg Lys Thr Thr 35
 40 25 30 35
 cag gat gtc aaa aaa aga aga gag gag gtc acc aag aaa cat cag tat 200
 Gln Asp Val Lys Lys Arg Arg Glu Glu Val Thr Lys Lys His Gln Tyr 50
 45 40 45 50
 gaa att agg aat tgt tgg cca cct gta tta tct ggg ggg atc agt cct 248
 Glu Ile Arg Asn Cys Trp Pro Pro Val Leu Ser Gly Gly Ile Ser Pro 65
 55 55 60 65
 tgc att atc att gaa aca cct cac aaa gaa ata gga aca agt gat ttc 296
 Cys Ile Ile Ile Glu Thr Pro His Lys Glu Ile Gly Thr Ser Asp Phe 80
 50 70 75 80
 tcc aga ttt aca aat tac aga ttt aaa aat ctt ttt att aat cct tca 344
 Ser Arg Phe Thr Asn Tyr Arg Phe Lys Asn Leu Phe Ile Asn Pro Ser 100
 55 85 90 95 100
 cct ttg cct gat tta agc tgg gga tgt tca aaa gaa gtc tgg cta aac 392
 Pro Leu Pro Asp Leu Ser Trp Gly Cys Ser Lys Glu Val Trp Leu Asn 115
 60 105 110 115
 atg tta aaa aag gag agc aga tat gtt cat gac aaa cat ttt gaa gtt 440
 Met Leu Lys Lys Glu Ser Arg Tyr Val His Asp Lys His Phe Glu Val 130
 120 125 130

SUBSTITUTE SHEET (rule 26)

5	ctg cat tct gac ttg gaa cca cag atg agg tcc ata ctt cta gac tgg 488 Leu His Ser Asp Leu Glu Pro Gln Met Arg Ser Ile Leu Leu Asp Trp 135 140 145
10	ctt tta gag gta tgt gaa gta tac aca ctt cat agg gaa aca ttt tat 536 Leu Leu Glu Val Cys Glu Val Tyr Thr Leu His Arg Glu Thr Phe Tyr 150 155 160
15	ctt gca caa gac ttt ttt gat aga ttt atg ttg aca caa aag gat ata 584 Leu Ala Gln Asp Phe Phe Asp Arg Phe Met Leu Thr Gln Lys Asp Ile 165 170 175 180
20	aat aaa aat atg ctt caa ctc att gga att acc tca tta ttc att gct 632 Asn Lys Asn Met Leu Gln Leu Ile Gly Ile Thr Ser Leu Phe Ile Ala 185 190 195
25	tcc aaa ctt gag gaa atc tat gct cct aaa ctc caa gag ttt gct tac 680 Ser Lys Leu Glu Glu Ile Tyr Ala Pro Lys Leu Gln Glu Phe Ala Tyr 200 205 210
30	gtc act gat ggt gct tgc agt gaa gaa gat atc tta agg atg gaa ctc 728 Val Thr Asp Gly Ala Cys Ser Glu Glu Asp Ile Leu Arg Met Glu Leu 215 220 225
35	att ata tta aag gct tta aaa tgg gaa ctt tgt cct gta aca atc atc 776 Ile Ile Leu Lys Ala Leu Lys Trp Glu Leu Cys Pro Val Thr Ile Ile 230 235 240
40	tcc tgg cta aat ctc ttt ctc caa gtt gat gct ctt aaa gat gct cct 824 Ser Trp Leu Asn Leu Phe Leu Gln Val Asp Ala Leu Lys Asp Ala Pro 245 250 255 260
45	aaa gtt ctt cta cct cag tat tct cag gaa aca ttc att caa ata gct 872 Lys Val Leu Leu Pro Gln Tyr Ser Gln Glu Thr Phe Ile Gln Ile Ala 265 270 275
50	cag ctt tta gat ctg tgt att cta gcc att gat tca tta gag ttc cag 920 Gln Leu Leu Asp Leu Cys Ile Leu Ala Ile Asp Ser Leu Glu Phe Gln 280 285 290
55	tac aga ata ctg act gct gct gcc ttg tgc cat ttt acc tcc att gaa 968 Tyr Arg Ile Leu Thr Ala Ala Ala Leu Cys His Phe Thr Ser Ile Glu 295 300 305
60	gtg gtt aag aaa gcc tca ggt ttg gag tgg gac agt att tca gaa tgt 1016 Val Val Lys Lys Ala Ser Gly Leu Glu Trp Asp Ser Ile Ser Glu Cys 310 315 320
65	gta gat tgg atg gta cct ttt gtc aat gta gta aaa agt act agt cca 1064 Val Asp Trp Met Val Pro Phe Val Asn Val Val Lys Ser Thr Ser Pro 325 330 335 340
70	gtg aag ctg aag act ttt aag aag att cct atg gaa gac aga cat aat 1112 Val Lys Leu Lys Thr Phe Lys Lys Ile Pro Met Glu Asp Arg His Asn 345 350 355
75	atc cag aca cat aca aac tat ttg gct atg ctg gag gaa gta aat tac 1160 Ile Gln Thr His Thr Asn Tyr Leu Ala Met Leu Glu Glu Val Asn Tyr 360 365 370

5 ata aac acc ttc aga aaa ggg gga cag ttg tca cca gtg tgc aat gga 1208
 Ile Asn Thr Phe Arg Lys Gly Gly Gln Leu Ser Pro Val Cys Asn Gly
 375 380 385

10 ggc att atg aca cca ccg aag agc act gaa aaa cca cca gga aaa cac 1256
 Gly Ile Met Thr Pro Pro Lys Ser Thr Glu Lys Pro Pro Gly Lys His
 390 395 400

15 taaagaagat aactaagcaa acaagttgga attcaccaag attgggtaga actggtatca 1316
 ctgaactact aaagttttac agaaagtagt gctgtgattg attgccctag ccaattcaca 1376
 agttacactg ccattctgat tttaaaactt acaattggca ctaaagaata catttaatta 1436
 tttcctatgt tagctgttaa agaaacagca ggacttggtt acaaagatgt cttcattccc 1496

20 aaggttactg gatagaagcc aaccacagtc tataccatag caatgttttt cctttaatcc 1556
 agtgttactg tgtttatctt gataaactag gaattttgtc actggagttt tggactggat 1616
 aagtgtacc ttaaagggtg tactaagtga tacagtactt tgaatctagt tgtagattc 1676

25 tcaaaattcc tacactcttg actagtgc aa tttggttctt gaaaattaaa tttaaacttg 1736
 tttaciaaagg tttagttttg taataagggt actaatttat ctatagctgc tatagcaagc 1796

30 tattataaaa cttgaatttc tacaaatggt gaaattta at gtttttttaa ctagtttatt 1856
 tgccttgcca taacacattt tttactaat aaggccttaga tgaacatggt gttcaacctg 1916

35 tgctctaaac agtgggagta ccaaagaaat tataaacaag ataatgctg tggctccttc 1976
 ctaactgggg ctttcttgac atgtagggtg cttggtaata acctttttgt atatcacaat 2036
 ttgggtgaaa aacttaagta ccctttcaaa ctatttatat gaggaagtca ctttactact 2096

40 ctaagatata cctaaggat tttttttttt aatttagtgt gactaaggct ttatttatgt 2156
 ttgtgaaact gttaagggtc tttctaaatt cctccattgt gagataagga cagtgtcaaa 2216

45 gtgataaagc ttaacacttg acctaaactt ctattttctt aaggaagaag agtattaaat 2276
 atatactgac tcctagaaat ctatttatta aaaaaagaca tgaaaacttg ctgtacatag 2336
 gctagctatt tctaaatatt ttaaattagc ttttctaaaa aaaaaatcca gcctcataaa 2396

50 gtagattaga aaactagatt gctagtttat tttgttatca gatatgtgaa tctcttctcc 2456
 ctttgaagaa actatacatt tattgttacg gtatgaagtc ttctgtatag tttgttttta 2516

55 aactaatatt tgtttcagta tttgtctgaa aaagaaaaca ccactaattg tgtacatatg 2576
 tattatataa acttaacctt ttaatactgt ttatttttag cccatgttta aaaaataaaa 2636
 gttaaaaaaa ttaactgct aaaaaaaaaa aaaaaaaagt gcggccgc 2684

60 <210> 54
 <211> 404
 <212> PRT

71

5 <213> Unknown
 <400> 54
 Met Ser Arg Arg Ser Ser Arg Leu Gln Ala Lys Gln Gln Pro Gln Pro
 1 5 10 15
 10 Ser Gln Thr Glu Ser Pro Gln Glu Ala Gln Ile Ile Gln Ala Lys Lys
 20 25 30
 Arg Lys Thr Thr Gln Asp Val Lys Lys Arg Arg Glu Glu Val Thr Lys
 35 40 45
 15 Lys His Gln Tyr Glu Ile Arg Asn Cys Trp Pro Pro Val Leu Ser Gly
 50 55 60
 20 Gly Ile Ser Pro Cys Ile Ile Ile Glu Thr Pro His Lys Glu Ile Gly
 65 70 75 80
 Thr Ser Asp Phe Ser Arg Phe Thr Asn Tyr Arg Phe Lys Asn Leu Phe
 85 90 95
 25 Ile Asn Pro Ser Pro Leu Pro Asp Leu Ser Trp Gly Cys Ser Lys Glu
 100 105 110
 Val Trp Leu Asn Met Leu Lys Lys Glu Ser Arg Tyr Val His Asp Lys
 115 120 125
 30 His Phe Glu Val Leu His Ser Asp Leu Glu Pro Gln Met Arg Ser Ile
 130 135 140
 35 Leu Leu Asp Trp Leu Leu Glu Val Cys Glu Val Tyr Thr Leu His Arg
 145 150 155 160
 Glu Thr Phe Tyr Leu Ala Gln Asp Phe Phe Asp Arg Phe Met Leu Thr
 165 170 175
 40 Gln Lys Asp Ile Asn Lys Asn Met Leu Gln Leu Ile Gly Ile Thr Ser
 180 185 190
 Leu Phe Ile Ala Ser Lys Leu Glu Glu Ile Tyr Ala Pro Lys Leu Gln
 195 200 205
 45 Glu Phe Ala Tyr Val Thr Asp Gly Ala Cys Ser Glu Glu Asp Ile Leu
 210 215 220
 50 Arg Met Glu Leu Ile Ile Leu Lys Ala Leu Lys Trp Glu Leu Cys Pro
 225 230 235 240
 Val Thr Ile Ile Ser Trp Leu Asn Leu Phe Leu Gln Val Asp Ala Leu
 245 250 255
 55 Lys Asp Ala Pro Lys Val Leu Leu Pro Gln Tyr Ser Gln Glu Thr Phe
 260 265 270
 Ile Gln Ile Ala Gln Leu Leu Asp Leu Cys Ile Leu Ala Ile Asp Ser
 275 280 285
 60 Leu Glu Phe Gln Tyr Arg Ile Leu Thr Ala Ala Ala Leu Cys His Phe
 290 295 300

SUBSTITUTE SHEET (rule 26)

72

5 Thr Ser Ile Glu Val Val Lys Lys Ala Ser Gly Leu Glu Trp Asp Ser
 305 310 315 320
 Ile Ser Glu Cys Val Asp Trp Met Val Pro Phe Val Asn Val Val Lys
 325 330 335
 10 Ser Thr Ser Pro Val Lys Leu Lys Thr Phe Lys Lys Ile Pro Met Glu
 340 345 350
 Asp Arg His Asn Ile Gln Thr His Thr Asn Tyr Leu Ala Met Leu Glu
 355 360 365
 15 Glu Val Asn Tyr Ile Asn Thr Phe Arg Lys Gly Gly Gln Leu Ser Pro
 370 375 380
 20 Val Cys Asn Gly Gly Ile Met Thr Pro Pro Lys Ser Thr Glu Lys Pro
 385 390 395 400
 Pro Gly Lys His
 25
 <210> 55
 <211> 1764
 <212> DNA
 <213> Unknown
 30
 <220>
 <223> Description of Unknown Organism:primate
 <220>
 35 <221> CDS
 <222> (26)..(1210)
 <400> 55
 40 ccgggatgcg aaggagcggg acacc atg aag gag gac ggc ggc gcg gag ttc 52
 Met Lys Glu Asp Gly Gly Ala Glu Phe
 1 5
 tcg gct cgc tcc agg aag agg aag gca aac gtg acc gtt ttt ttg cag 100
 Ser Ala Arg Ser Arg Lys Arg Lys Ala Asn Val Thr Val Phe Leu Gln
 45 10 15 20 25
 gat cca gat gaa gaa atg gcc aaa atc gac agg acg gcg agg gac cag 148
 Asp Pro Asp Glu Glu Met Ala Lys Ile Asp Arg Thr Ala Arg Asp Gln
 30 35 40
 50 tgt ggg agc cag cct tgg gac aat aat gca gtc tgt gca gac ccc tgc 196
 Cys Gly Ser Gln Pro Trp Asp Asn Asn Ala Val Cys Ala Asp Pro Cys
 45 50 55
 55 tcc ctg atc ccc aca cct gac aaa gaa gat gat gac cgg gtt tac cca 244
 Ser Leu Ile Pro Thr Pro Asp Lys Glu Asp Asp Asp Arg Val Tyr Pro
 60 60 65 70
 aac tca acg tgc aag cct cgg att att gca cca tcc aga ggc tcc ccg 292
 Asn Ser Thr Cys Lys Pro Arg Ile Ile Ala Pro Ser Arg Gly Ser Pro
 75 80 85
 ctg cct gta ctg agc tgg gca aat aga gag gaa gtc tgg aaa atc atg 340

SUBSTITUTE SHEET (rule 26)

5	Leu	Pro	Val	Leu	Ser	Trp	Ala	Asn	Arg	Glu	Glu	Val	Trp	Lys	Ile	Met	
	90					95					100					105	
	tta	aac	aag	gaa	aag	aca	tac	tta	agg	gat	cag	cac	ttt	ctt	gag	caa	388
	Leu	Asn	Lys	Glu	Lys	Thr	Tyr	Leu	Arg	Asp	Gln	His	Phe	Leu	Glu	Gln	
10					110					115					120		
	cac	cct	ctt	ctg	cag	cca	aaa	atg	cga	gca	att	ctt	ctg	gat	tgg	tta	436
	His	Pro	Leu	Leu	Gln	Pro	Lys	Met	Arg	Ala	Ile	Leu	Leu	Asp	Trp	Leu	
				125					130					135			
15	atg	gag	gtg	tgt	gaa	gtc	tat	aaa	ctt	cac	agg	gag	acc	ttt	tac	ttg	484
	Met	Glu	Val	Cys	Glu	Val	Tyr	Lys	Leu	His	Arg	Glu	Thr	Phe	Tyr	Leu	
			140					145					150				
20	gca	caa	gat	ttc	ttt	gac	cgg	tat	atg	gcg	aca	caa	gaa	aat	gtt	gta	532
	Ala	Gln	Asp	Phe	Phe	Asp	Arg	Tyr	Met	Ala	Thr	Gln	Glu	Asn	Val	Val	
		155					160					165					
25	aaa	act	ctt	tta	cag	ctt	att	ggg	att	tca	tct	tta	ttt	att	gca	gcc	580
	Lys	Thr	Leu	Leu	Gln	Leu	Ile	Gly	Ile	Ser	Ser	Leu	Phe	Ile	Ala	Ala	
	170					175					180					185	
	aaa	ctt	gag	gaa	atc	tat	cct	cca	aag	ttg	cac	cag	ttt	gcg	tat	gtg	628
	Lys	Leu	Glu	Glu	Ile	Tyr	Pro	Pro	Lys	Leu	His	Gln	Phe	Ala	Tyr	Val	
					190					195					200		
30	aca	gat	gga	gct	tgt	tca	gga	gat	gaa	att	ctc	acc	atg	gaa	tta	atg	676
	Thr	Asp	Gly	Ala	Cys	Ser	Gly	Asp	Glu	Ile	Leu	Thr	Met	Glu	Leu	Met	
				205					210					215			
35	att	atg	aag	gcc	ctt	aag	tgg	cgt	tta	agt	ccc	ctg	act	att	gtg	tcc	724
	Ile	Met	Lys	Ala	Leu	Lys	Trp	Arg	Leu	Ser	Pro	Leu	Thr	Ile	Val	Ser	
			220					225					230				
40	tgg	ctg	aat	gta	tac	atg	cag	gtt	gca	tat	cta	aat	gac	tta	cat	gaa	772
	Trp	Leu	Asn	Val	Tyr	Met	Gln	Val	Ala	Tyr	Leu	Asn	Asp	Leu	His	Glu	
		235					240					245					
45	gtg	cta	ctg	ccg	cag	tat	ccc	cag	caa	atc	ttt	ata	cag	att	gca	gag	820
	Val	Leu	Leu	Pro	Gln	Tyr	Pro	Gln	Gln	Ile	Phe	Ile	Gln	Ile	Ala	Glu	
	250					255					260					265	
	ctg	ttg	gat	ctc	tgt	gtc	ctg	gat	gtt	gac	tgc	ctt	gaa	ttt	cct	tat	868
	Leu	Leu	Asp	Leu	Cys	Val	Leu	Asp	Val	Asp	Cys	Leu	Glu	Phe	Pro	Tyr	
					270					275					280		
50	ggt	ata	ctt	gct	gct	tcg	gcc	ttg	tat	cat	ttc	tcg	tca	tct	gaa	ttg	916
	Gly	Ile	Leu	Ala	Ala	Ser	Ala	Leu	Tyr	His	Phe	Ser	Ser	Ser	Glu	Leu	
				285					290					295			
55	atg	caa	aag	gtt	tca	ggg	tat	cag	tgg	tgc	gac	ata	gag	aac	tgt	gtc	964
	Met	Gln	Lys	Val	Ser	Gly	Tyr	Gln	Trp	Cys	Asp	Ile	Glu	Asn	Cys	Val	
			300					305					310				
60	aag	tgg	atg	gtt	cca	ttt	gcc	atg	gtt	ata	agg	gag	acg	ggg	agc	tca	1012
	Lys	Trp	Met	Val	Pro	Phe	Ala	Met	Val	Ile	Arg	Glu	Thr	Gly	Ser	Ser	
		315					320					325					
	aaa	ctg	aag	cac	ttc	agg	ggc	gtc	gct	gat	gaa	gat	gca	cac	aac	ata	1060

SUBSTITUTE SHEET (rule 26)

5	Lys Leu Lys His Phe Arg Gly Val Ala Asp Glu Asp Ala His Asn Ile 330 335 340 345	
	cag acc cac aga gac agc ttg gat ttg ctg gac aaa gcc cga gca aag Gln Thr His Arg Asp Ser Leu Asp Leu Leu Asp Lys Ala Arg Ala Lys 350 355 360	1108
10	aaa gcc atg ttg tct gaa caa aat agg gct tct cct ctc ccc agt ggg Lys Ala Met Leu Ser Glu Gln Asn Arg Ala Ser Pro Leu Pro Ser Gly 365 370 375	1156
15	ctc ctc acc ccg cca cag agc ggt aag aag cag agc agc ggg ccg gaa Leu Leu Thr Pro Pro Gln Ser Gly Lys Lys Gln Ser Ser Gly Pro Glu 380 385 390	1204
20	atg gcg tgaccacccc atccttctcc accaaagaca gttgcgcgcg tgctccacgt Met Ala 395	1260
	tctcttctgt ctgttgcagc ggaggcgtgc gtttgctttt acagatatct gaatggaaga	1320
25	gtgtttcttc cacaacagaa gtatttctgt ggatggcatc aaacagggca aagtgttttt	1380
	tattgaatgc ttatagggttt tttttaaata agtgggtcaa gtacaccagc cacctccaga	1440
30	caccagtgcg tgctcccgat gctgctatgg aagggtgctac ttgacctaaag ggactcccac	1500
	aacaacaaaa gcttgaagct gtggaggcgc acggtggcgt ggctctcctc gcagggtgttc	1560
	tgggctccgt tgtaccaagt ggagcaggtg gttgcgggca agcgttgtgc agagcccata	1620
35	gccagctggg cagggggctg cctctccac attatcagtt gacagtgtac aatgcctttg	1680
	atgaactggt ttgtaagtgc tgctatatct atccattttt taataaagct aatactgttt	1740
40	ctttagagca cactggcggg tcgt	1764
	<210> 56 <211> 395 <212> PRT <213> Unknown	
45	<400> 56 Met Lys Glu Asp Gly Gly Ala Glu Phe Ser Ala Arg Ser Arg Lys Arg 1 5 10 15	
50	Lys Ala Asn Val Thr Val Phe Leu Gln Asp Pro Asp Glu Glu Met Ala 20 25 30	
	Lys Ile Asp Arg Thr Ala Arg Asp Gln Cys Gly Ser Gln Pro Trp Asp 35 40 45	
55	Asn Asn Ala Val Cys Ala Asp Pro Cys Ser Leu Ile Pro Thr Pro Asp 50 55 60	
60	Lys Glu Asp Asp Asp Arg Val Tyr Pro Asn Ser Thr Cys Lys Pro Arg 65 70 75 80	
	Ile Ile Ala Pro Ser Arg Gly Ser Pro Leu Pro Val Leu Ser Trp Ala 85 90 95	

75

5	Asn Arg Glu Glu Val Trp Lys Ile Met Leu Asn Lys Glu Lys Thr Tyr	100	105	110
	Leu Arg Asp Gln His Phe Leu Glu Gln His Pro Leu Leu Gln Pro Lys	115	120	125
10	Met Arg Ala Ile Leu Leu Asp Trp Leu Met Glu Val Cys Glu Val Tyr	130	135	140
	Lys Leu His Arg Glu Thr Phe Tyr Leu Ala Gln Asp Phe Phe Asp Arg	145	150	155
15	Tyr Met Ala Thr Gln Glu Asn Val Val Lys Thr Leu Leu Gln Leu Ile	165	170	175
	Gly Ile Ser Ser Leu Phe Ile Ala Ala Lys Leu Glu Glu Ile Tyr Pro	180	185	190
20	Pro Lys Leu His Gln Phe Ala Tyr Val Thr Asp Gly Ala Cys Ser Gly	195	200	205
	Asp Glu Ile Leu Thr Met Glu Leu Met Ile Met Lys Ala Leu Lys Trp	210	215	220
	Arg Leu Ser Pro Leu Thr Ile Val Ser Trp Leu Asn Val Tyr Met Gln	225	230	235
30	Val Ala Tyr Leu Asn Asp Leu His Glu Val Leu Leu Pro Gln Tyr Pro	245	250	255
	Gln Gln Ile Phe Ile Gln Ile Ala Glu Leu Leu Asp Leu Cys Val Leu	260	265	270
35	Asp Val Asp Cys Leu Glu Phe Pro Tyr Gly Ile Leu Ala Ala Ser Ala	275	280	285
	Leu Tyr His Phe Ser Ser Ser Glu Leu Met Gln Lys Val Ser Gly Tyr	290	295	300
	Gln Trp Cys Asp Ile Glu Asn Cys Val Lys Trp Met Val Pro Phe Ala	305	310	315
45	Met Val Ile Arg Glu Thr Gly Ser Ser Lys Leu Lys His Phe Arg Gly	325	330	335
	Val Ala Asp Glu Asp Ala His Asn Ile Gln Thr His Arg Asp Ser Leu	340	345	350
50	Asp Leu Leu Asp Lys Ala Arg Ala Lys Lys Ala Met Leu Ser Glu Gln	355	360	365
	Asn Arg Ala Ser Pro Leu Pro Ser Gly Leu Leu Thr Pro Pro Gln Ser	370	375	380
	Gly Lys Lys Gln Ser Ser Gly Pro Glu Met Ala	385	390	395
60				

SUBSTITUTE SHEET (rule 26)